

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: February 25, 2003, 16:09:53 ; Search time 43 Seconds
(without alignments)
10598.073 Million cell updates/sec
Title: US-09-762-021A-1
Perfect score: 3127
Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaataangataaa 1710
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Command line parameters:
-MODEL=frame+L2p.model -DEV=xlh
-Q=/Cyn2.1/USPTO.spool/US09762021/runat_24022003_153037_7535/app_query.fasta_1.1863
-DB=A_Geneseq_101002 -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEXT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021 8CGN_1_141 @runat_24022003_153037_7535 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2421	77.4	593	22	AAB64372 Amino acid sequenc
2	912	29.2	179	22	AAU87627 Novel central nerv
3	912	29.2	179	22	AAU17192 Novel signal trans
4	912	29.2	179	22	AAG74084 Human colon cancer
5	912	29.2	185	22	AAU87344 Novel central nerv
6	691.5	22.1	348	22	AAM92275 Human digestive sy
7	603	19.3	121	22	AAU87343 Novel central nerv
8	602	19.3	121	22	AAU17191 Novel signal trans
9	598	19.1	134	22	AAU87626 Novel central nerv
10	598	19.1	134	22	AAU17570 Novel signal trans
11	584	18.7	535	22	AAM25595 Human protein sequ
12	579.5	18.5	723	22	AAM93529 Human polypeptide,
13	561	17.9	569	22	AAU31985 Novel human secret
14	558.5	17.9	821	14	AAR35451 Mouse eps8. Mus m
15	482	15.4	112	22	AAG75249 Human colon cancer
16	452	14.5	549	22	ABG09715 Novel human diagno
17	403	12.9	778	22	ABG60520 Drosophila melanog
18	403	12.9	778	22	ABG66160 Drosophila melanog
19	341.5	10.9	714	22	ABB71668 Drosophila melanog
20	327.5	10.5	327	22	AAB93385 Human protein sequ
21	298	9.5	764	22	ABG09702 Novel human diagno
22	271	8.7	328	22	AAG73861 Human colon cancer
23	261	8.3	75	22	AAG77134 Human colon cancer
24	256.5	8.2	280	22	AAU87645 Novel central nerv
25	256.5	8.2	280	22	ABBI0961 Human ovarian and/
26	256.5	8.2	280	22	AAU96583 Human reproductive
27	256.5	8.2	280	22	AAU17560 Novel signal trans
28	256.5	8.2	280	23	ABP41618 Human ovarian anti
29	256.5	8.2	283	22	AAU87365 Novel central nerv
30	236.5	8.2	283	22	AAU17147 Novel signal trans
31	221.5	7.2	1040	22	ABG14734 Novel human diagno
32	200.5	6.4	1078	16	AAU1704 Collagen alpha 1 (
33	200.5	6.4	1466	22	ABBS0291 Collagen type III
34	200.5	6.4	1466	23	ABBS0747 Human Tumour Endot
35	200.5	6.4	1469	22	ABG15191 Novel human diagno
36	198	6.3	1466	22	AAE02537 Porcine alphasIII
37	195.5	6.3	1196	13	AAU28916 Type III procollag
38	194	6.2	686	22	ABBI1397 Human secreted pro
39	193.5	6.2	1466	22	AAE02533 Bovine alpha(III)
40	193.5	6.2	1466	22	AAE02534 Bovine alpha(III)
41	191	6.1	1078	21	AAU96125 Collagen type III
42	191	6.1	1078	23	ABBS0736 Collagen type III
43	191	6.1	1078	23	ABBS0628 Amino acid sequenc
44	191	6.1	1078	23	AAE16478 Human collagen alp
45	191	6.1	1963	22	ABBS62819 Drosophila melanog

ALIGNMENTS

RESULT 1

AAB64372
ID AAB64372 standard; Protein: 593 AA.
XX
AC AAB64372;
DT
XX 22-MAR-2001 (first entry)
DE Amino acid sequence of human intracellular signalling molecule INTRA4.

Human; intracellular signalling molecule; INTRA; immunosuppressive;
cytostatic; neuroprotective; nootropic; antiarteriosclerotic; cancer;
antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal;
antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS;
cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;
inflammatory disorder; Addison's disease; gastrointestinal disorder;
neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;
mental disorder; schizophrenia; anxiety.

OS Homo sapiens.
 PN WO200077040-A2.
 XX 21-DEC-2000.
 PD
 XX 16-JUN-2000; 2000WO-US16636.
 XX 16-JUN-1999; 99US-0139566.
 PR 17-AUG-1999; 99US-0149640.
 PR 09-NOV-1999; 99US-0164417.
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Yue H, Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;
 PI Azimzai Y, Yang J, Reddy R, Lu DAM;
 PI WPI; 2001-025334/03.
 DR N-PSDB; AAF32641.
 XX New human intracellular signaling molecules, useful for the diagnosis,
 PT prevention and treatment of cell proliferative, autoimmune, and
 PT inflammatory, neurological, gastrointestinal, reproductive, and
 PT developmental disorders -
 XX Claim 5; Page 115-116; 192pp; English.
 PS Sequences AAF32638 - AAF32689 represent cDNA encoding human
 CC intracellular signalling molecules INTRAL - INTRAS2, represented in
 CC AAF64369 - AAF64420. Modulators of the intracellular signalling molecules
 CC of the invention exhibit immunosuppressive; cytostatic; neuroprotective;
 CC neurotropic; antiarteriosclerotic; antiinflammatory; anti-HIV;
 CC antileptic; antibacterial; antifungal; antiviral; antiparasitic;
 CC antineoplastic; and antiparkinsonian activity. INTRA polypeptides, their
 CC agonists and antagonists are useful for the treatment of a condition
 CC associated with decreased or increased expression of functional INTRA.
 CC Disorders associated with abnormal INTRA expression or activity include
 CC cell proliferative disorders e.g. arteriosclerosis and cancers;
 CC autoimmune or inflammatory disorders e.g. Addison's disease and acquired
 CC immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,
 CC protozoal and helminthic infections; gastrointestinal disorders e.g.
 CC dysphagia and irritable bowel syndrome; neurological disorders e.g.
 CC epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob
 CC disease and mental disorders e.g. anxiety, schizophrenia and Tourette's
 CC disorder. Antibodies immuno specific for the INTRA proteins may also be
 CC useful in the diagnosis of the above disorders.
 XX Sequence 593 AA;
 SQ

Alignment Scores:
 Pred. No.: 1,74e-190 Length: 593
 Score: 2421.00 Matches: 457
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 77.42% Indels: 0
 DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAB64372 (1-593)

QY 2 GCAGAGGACTGAAGACCCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAAACCTT 61
 |||||
 DB 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluGluGlnArgPro 155
 |||||
 QY 62 CGACTTGGAGGCTTCAGCCAGCCAGGACAGATGAGGGGGCTGCTATGGAAAGCCG 121
 |||||
 DB 156 ArgLeuGlyGlyLeuGlnProSerGlnAspArgTrpArgGlyProAlaMetGluArgPro 175
 |||||
 QY 122 CTCCTATGAGGAGGACGACCTATCTGGAGCCGGGATCCCTCCAGAACAGCCACCCAG 181
 |||||
 DB 176 LeuProMetGluGlnAlaArgTyLeuGluProGlyIleProGluGlnProHisGln 195
 |||||
 QY 182 AGACCCCTAGACACAGCCTCCACCATCCCAAGGCCCTGCCAGCCACACCATGGCC 241
 |||||

DB 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
 |||||
 QY 242 CGAGAAACCAAGTGCCTTTACTCTGCCTCTCCAAAGCGGTCTCTTCCCGGAGGACCCA 301
 |||||
 DB 216 ArgGluProSerAlaPheThrLeuProProArgSerSerSerProGluAspPro 235
 |||||
 QY 302 GAGAGGACGAGGAAGTGTGAACCATGTCTAAGGACATTTGAGCTGTTCATGGGAAAG 361
 |||||
 DB 236 GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
 |||||
 QY 362 CTGGAGAAGCCCGCAGCAAGACCAGCAGAGAAATAATTTGGGAAAAAACACAGGAC 421
 |||||
 DB 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysPheGlyLysLysAsnLysAsp 275
 |||||
 QY 422 CAGGGAGGTCTCACCAGGCACAGTACATTGCTGCTCCAGAAAGATCAAGTACAGCTTC 481
 |||||
 DB 276 GlnGlyGlyLeuThrGlnAlaGlnIleAspCysPheGlnLysIleLysIleSerPhe 295
 |||||
 QY 482 AACCTCTCTGGGAAGGCTGCCACCTGGCTGAAGGACAAAGTGCCTCGAGCTCGTACAC 541
 |||||
 DB 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
 |||||
 QY 542 ATCTCTTCAAGTCCCTGAACCTTCATCTCGGCAGGTGCCCTGAGGCTGGCTAGCAGCC 601
 |||||
 DB 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAla 335
 |||||
 QY 602 CAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCCTACAGTCTCTCTAAGC 661
 |||||
 DB 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeuSer 355
 |||||
 QY 662 CCACCTGAGAGTAACCTTTGGATGGGGTTGGGCCACCTGGACCACTAGCCGGGGCCGAC 721
 |||||
 DB 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaThrThrSerArgAlaAsp 375
 |||||
 QY 722 TGGACAGGCGATGAGCCCTGCCCTACCAACCCACCATTTCTCAGATGACTGGCACTTCCA 781
 |||||
 DB 376 TrpThrGlyAspGluProLeuProTyTrpGlnProThrPheSerAspAspTrpGlnLeu 395
 |||||
 QY 782 GAGCCCTCCAGCCCAACACCCCTTAGGATACCAAGACCCCTGTTTCCCTCGCGGGGAAAT 841
 |||||
 DB 396 GluProSerSerGlnAlaProLeuGlyTyTrpGlnAspProValSerLeuArgArgGlySer 415
 |||||
 QY 842 CATAGTTAGGAGACCTCACAATTTCTCAGGAGAAGACACACCAACCATGACCTCAG 901
 |||||
 DB 416 HisArgLeuGlySerThrSerHisPheProGlnGluLysThrHisAsnHisAspProGln 435
 |||||
 QY 902 CTTGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATG 961
 |||||
 DB 436 ProGlyAspProAsnSerArgProSerProLysProAlaGlnProAlaLeuLysMet 455
 |||||
 QY 962 CAAGTCTTGTACAGTTTGAAGCTAGGAACCCACGGGAACCTGACTGTGTCAGGAGGAG 1021
 |||||
 DB 456 GlnValLeuTyTrpGluPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGlu 475
 |||||
 QY 1022 AAGCTGGAGTTCTTGACACACAGCAGCGGTGGTGGTGAAGATGAGCGGGGACGG 1081
 |||||
 DB 476 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 495
 |||||
 QY 1082 AGCGGTACATTCACAGCAACATCCTGGAGCCCTACAGCCGGGACCCCTGGAGCCCGAG 1141
 |||||
 DB 496 SerGlyTyTrpIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 515
 |||||
 QY 1142 GGCAGCTCACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCTCGAAGAGGTCACA 1201
 |||||
 DB 516 GlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluValThr 535
 |||||
 QY 1202 GACTGTGCTCGAGCGAGAACCTTCTCCTACCTGCCACCGGTGAGGACACTTGGGTCCCTGACG 1261
 |||||
 DB 536 AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 555
 |||||
 QY 1262 GSGAGCCAGCTACTCTGCATAGACCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCC 1321
 |||||
 DB 556 GlySerGlnLeuLeuArgIleArgProGlyGluGlnMetLeuCysProGlnGluAla 575
 |||||

Qy	1322	CCAGAAATCTGTCCTCCGGCTGGAGCTGTGACAGGATGCTGGGATTAAGCCCT	1375
Db	576	ProArgIleLeuSerArgLeuGluAlaValArgMetLeuGlyIleSerPro	593
RESULT 2			
AAU87627			
ID	AAU87627	standard; Protein; 179 AA.	
AC	AAU87627;		
XX	XX		
DT	05-JUN-2002	(first entry)	
DE	Novel central nervous system protein #537.		
XX		Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;	
KW		hyperproliferative disorder; neoplasia; cardiovascular disorder;	
KW		cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;	
KW		nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;	
KW		acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;	
KW		adenocarcinoma; reproductive system disorder; testicular feminisation;	
KW		endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;	
KW		respiratory disorder; renal disorder; kidney failure; blood disorder;	
KW		myocardial infarction; wound healing; cell proliferation; skin aging;	
KW		food additive; food preservative; gene therapy.	
XX			
OS		homo sapiens.	
XX			
PN	WO200155318-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01332.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225274.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226686.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 20		

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SW;
XX
XX DR WPI; 2001-581633/65.
XX N-PSDB; ABK43957.
XX
XX PT New isolated nucleic acid encoding a protein for diagnosing,
XX PT preventing, treating or ameliorating medical conditions and used as
XX PT food additives or preservatives -
XX
XX PS Claim 9; SEQ ID No 1145; 837pp; English.
XX
XX CC The invention describes an isolated nucleic acid molecule (I) encoding a
XX CC novel central nervous system protein. (I) and polypeptides (III) encoded
XX CC by (I), are used to treat a medical conditions and in diagnosis of a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX CC anglogenesis, nervous system disorders e.g. Alzheimer's disease and
XX CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
XX CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
XX CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
XX CC adenocarcinomas and irritable bowel syndrome, reproductive system
XX CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
XX CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
XX CC leukaemia, disorders involving neovascularisation e.g. malignancies,
XX CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
XX CC acute kidney failure and blood related disorders e.g. myocardial
XX CC infarction. The polypeptides can also be used to aid wound healing and
XX CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX CC maintain organs before transplantation, for supporting cell culture of
XX CC primary tissues, to regenerate tissues and in chemotaxis. The
XX CC polypeptides can also be used as a food additive or preservative to
XX CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:
Pred. No.: 1 5e-66 Length: 179
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 22 Gaps: 1

US-09-762-021A-1 (1-1710) x AAU87627 (1-179)
QY 796 AGCACCTTAGGATACACAGGCCCTGTTTCCTTCGGGGGAGCTCATAGTTAGGAG 855

|||||
1 SerThrLeuArgIleProGlyProCysPhePro----- 11
Db
QY 856 CACCTCACACTTTCCTCAGGAGAGACACACACCATGACCTCAGCTGGGACCCCAA 915
Db
12 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 26
QY 916 CTCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTGTACGA 975
Db
26 nSerArgProSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyG1 46
QY 976 GTTTGAAGCTAGGAACCCACGGGAACCTGACTGTGCTCCAGGAGAGAAAGCTCGAGTTCT 1035
Db
46 uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluVal 66
QY 1036 GGACCACAGCAAGCGGTGGTGGTGAAGAATGAGCGGGAGCGGCGGTACATTCC 1095
Db
66 uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyIlePr 86
QY 1096 AAGCAACATCTCGAGCCCTACAGCGGGGACCCCTGGGACCCAGGCGCCAGTACCCCTC 1155
Db
86 oSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 106
QY 1156 TCGGTTCCCAATGCTTCGACTTAGCTCGAGCCCTGAAGAGGTTCACAGACTGCCTGCAGGC 1215
Db
106 rArgValProMetLeuArgLeuSerSerArgProGluGluValThrAspTrpLeuGlnAl 126
QY 1216 AGAGAACTTCTCCACTGCCAGCGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACT 1275
Db
126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
QY 1276 TCGATAAGACTCGGGAGCTACAGATGCTATGTCCAGGAGGCCCCACCAATCCTCTGC 1335
Db
146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
QY 1336 CCGCTGGAGCTGTGAGAAGGATGCTCGGGATAAGCCCT 1375
Db
166 rArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 179
RESULT 3
AAU17192
ID AAU17192 standard; Protein: 179 AA.
AC AAU17192;
XX
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID 757.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
XX acquired immune deficiency syndrome.
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229487.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

WPI; 2001-465460/50.

N-PSDB; AAS27109.

Novel polypeptides useful for diagnosing, treating, preventing and/or
prognosing disorders related to the proteins, including cancers, immune
disorders and neuronal disorders

Claim 1; SEQ ID No 757; 880pp; English.

The invention relates to novel isolated polypeptides (1), and

CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Alignment Scores:
Pred. No.: 1.5e-66 Length: 179
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 22 Gaps: 1

US-09-762-021A-1 (1-1710) x AAU17192 (1-179)

QY 796 AGCACCTTAGCATACAGGACCCCTTTCCCTCGCGGGGAAGTCATAGTTAGGGAG 855
DB 1 SerThrLeuArgIleProGlyProCysPhePro----- 11
QY 856 CACCTCACACTTTCCTCAGGAGAACACACACCAACCATGACCTCAGCCCTGGGACCCCAA 915
DB 12 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 26
QY 916 CTCAGGCCCTCAGCCCAACCTGCTCCAGCAGCCCTGAAATGCAAGTCTTGACGA 975
DB 26 nSerArgProSerSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrgl 46
QY 976 GTTTGAAGCTAGAACCCACGGAAGTACTGTGTGTCAGGAGAGAGCTGAGGTCT 1035
DB 46 uPheGluAlaArgAsnProArgGluLeuThrValGlnGlnGlyLysLeuGluVal 66
QY 1036 GGACACAGCAAGCGGTGTGCTGGTGAAGAAATGAGCGGACGAGCGGTACATTC 1095
DB 66 uAspHisSerLysArgTyrTrpLeuValLysAsnGluAlaGlyArgSerGlyTyrIlePr 86
QY 1096 AGCACATCTCGGACCCCTCAGCCGGGGACCCCTGGGACCCAGGCGCATCACCCTC 1155
DB 86 oSerAsnIleLeuGluProGlnProGlyThrProGlyThrGlnGlnSerProSe 106
QY 1156 TCGGGTTCCAATGCTCGACTAGCTCGAGGCTGAAGAGTCAAGAGTGCAGCTGCAGGC 1215
DB 106 rArgValProMetLeuArgLeuSerSerArgProGluGluValThrAspTrpLeuGlnAl 126
QY 1216 AGAAATCTTCTCCACTGCCCGGTGAGGACACTTGGGTCCCTCAGCGGGAGCCAGCTACT 1275
DB 126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
QY 1276 TGCATAGACCTCGGGGACTACAGTGTATGTCCACAGGAGGCCCCACGAATCTGTCT 1335
DB 146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
QY 1336 CGGGCTGGAGGTGTGCAGAGGATGCTGGGGATAAGCCCT 1375
DB 166 rArgLeuGluAlaValArgMetLeuGlyIleSerPro 179

RESULT 4
AAG74084

ID AAG74084 standard; Protein; 179 AA.
XX
AC AAG74084;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:4848.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH33515.
XX

XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
XX
XX Claim 11; Page 6631-6632; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
XX cancer-associated nucleic acid molecules (N) and proteins (P), where
XX the proteins are collectively known as colon cancer antigens. The colon
XX cancer antigens have cytostatic activity and can be used in gene
XX therapy and vaccine production. N and P may be used in the prevention,
XX diagnosis and treatment of diseases associated with inappropriate P
XX expression. For example, N and P may be used to treat disorders
XX associated with decreased expression by rectifying mutations or deletions
XX in a patient's genome that affect the activity of P by expressing
XX inactive proteins or to supplement the patients own production of P.
XX Additionally, N may be used to produce the colon cancer-associated Ps,
XX by inserting the nucleic acids into a host cell and culturing the cell,
XX to express the proteins. N and P can be used in the prevention, diagnosis
XX and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
XX and AAH37789 represent sequences used in the exemplification of the
XX present invention.
XX N.B. Pages 666 to 682 and page 7053 of the sequence listing were
XX missing at time of publication, meaning no sequences are present for
XX SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 179 AA:

Alignment Scores:
Pred. No.: 1.5e-66 Length: 179
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 22 Gaps: 1

US-09-762-021A-1 (1-1710) x AAG74084 (1-179)

QY 796 AGCACCTTAGCATACAGGACCCCTTTCCCTCGCGGGGAAGTCATAGTTAGGGAG 855
DB 1 SerThrLeuArgIleProGlyProCysPhePro----- 11
QY 856 CACCTCACACTTTCCTCAGGAGAACACACACCAACCATGACCTCAGCCCTGGGACCCCAA 915

Db	12	Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs	26	PR	28-JUN-2000;	2000US-0214886.
Qy	916	CTCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAATGCAAGTCCTGTACGA	975	PR	30-JUN-2000;	2000US-0215135.
Db	26	nSerArgProSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrGI	46	PR	07-JUL-2000;	2000US-0216647.
Qy	976	GTTTGAAGCTAGGACCCACCGGAACTGCTGTGTCAGGAGAGAGCTGGAGGTTCT	1035	PR	07-JUL-2000;	2000US-0216880.
Db	46	uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluValLe	66	PR	11-JUL-2000;	2000US-0217487.
Qy	1036	GGACACAGCAAGCGGTGGCTGGTGGAAGAATGAGCGCGGAGCGGCTACATTCC	1095	PR	14-JUL-2000;	2000US-0217496.
Db	66	uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyrIlePr	86	PR	14-JUL-2000;	2000US-0218290.
Qy	1096	AAGCAACATCTGAGCCCTACAGCCGGGACCCCTGGGACCCAGCGCCAGTCAACCTC	1155	PR	26-JUL-2000;	2000US-0220963.
Db	86	oSerAsnIleuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProse	106	PR	26-JUL-2000;	2000US-0220964.
Qy	1156	TCGGGTTCCAAATGCTTCGACTTAGCTCGAGGCCTGAAGAGTGCACAGCTGGCTGCAGGC	1215	PR	14-AUG-2000;	2000US-0224518.
Db	106	rArgValProMetLeuArgLeuSerSerArgProGluGluValThrAspTrpLeuGlnAl	126	PR	14-AUG-2000;	2000US-0224519.
Qy	1216	AGAGAACTTCTCCACTGCCAGCGTGAGACACTTGGGTCCCTGACGGGAGCCAGCTACT	1275	PR	14-AUG-2000;	2000US-0225213.
Db	126	aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe	146	PR	14-AUG-2000;	2000US-0225214.
Qy	1276	TCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGAGGCGCCACCAATCCTGTC	1335	PR	14-AUG-2000;	2000US-0225267.
Db	146	uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe	166	PR	14-AUG-2000;	2000US-0225270.
Qy	1336	CCGCTGGAGGCTGTCAGAGGATGCTGGGGATAAGCCCT	1375	PR	14-AUG-2000;	2000US-0225276.
Db	166	rArgLeuGluAlaValArgMetLeuGlyIleSerPro	179	PR	14-AUG-2000;	2000US-0225277.
RESULT 5					2000US-0225278.	
AAU87344					2000US-0225279.	
ID	AAU87344 standard; Protein: 185 AA.				2000US-0225280.	
XX					2000US-0225281.	
AC	AAU87344;				2000US-0225282.	
XX					2000US-0225283.	
XX	05-JUN-2002 (first entry)				2000US-0225284.	
DT	Novel central nervous system protein #254.				2000US-0225285.	
DE	Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;				2000US-0225286.	
KW	hyperproliferative disorder; neoplasm; cardiovascular disorder;				2000US-0225287.	
KW	cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;				2000US-0225288.	
KW	nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;				2000US-0225289.	
KW	acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;				2000US-0225290.	
KW	adenocarcinoma; reproductive system disorder; testicular feminisation;				2000US-0225291.	
KW	endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;				2000US-0225292.	
KW	respiratory disorder; renal disorder; kidney failure; blood disorder;				2000US-0225293.	
KW	myocardial infarction; wound healing; cell proliferation; skin aging;				2000US-0225294.	
KW	food additive; food preservative; gene therapy.				2000US-0225295.	
OS	Homo sapiens.				2000US-0225296.	
XX					2000US-0225297.	
XX	WO200155318-A2.				2000US-0225298.	
PN					2000US-0225299.	
XX	02-AUG-2001.				2000US-0225300.	
PD					2000US-0225301.	
XX	17-JAN-2001; 2001WO-US01332.				2000US-0225302.	
PF					2000US-0225303.	
XX	31-JAN-2000; 2000US-0179065.				2000US-0225304.	
PR	04-FEB-2000; 2000US-0180628.				2000US-0225305.	
PR	24-FEB-2000; 2000US-0184664.				2000US-0225306.	
PR	02-MAR-2000; 2000US-0186350.				2000US-0225307.	
PR	16-MAR-2000; 2000US-0189874.				2000US-0225308.	
PR	17-MAR-2000; 2000US-0190076.				2000US-0225309.	
PR	18-APR-2000; 2000US-0198123.				2000US-0225310.	
PR	19-MAY-2000; 2000US-0205515.				2000US-0225311.	
PR	07-JUN-2000; 2000US-0209467.				2000US-0225312.	

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

XX PA Rosen CA, Barash SC, Ruben SW;
XX WPI; 2001-581633/65.
DR N-PSDB; ABK43674.

XX New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -

XX Claim 9; SEQ ID No 862; 837pp; English.

XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:

Pred. No.: 1,52e-66 Length: 185
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 22 Gaps: 1

US-09-762-021a-1 (1-1710) x AAU87344 (1-185)

Qy 796 AGCACCCCTTAGGATACAGGACCCCTGTTTCCTTCGCGGGGAGTCATAGGTTAGGAG 855

Db 7 SerThrLeuArgIleProGlyProCysPhePro----- 17

Qy 856 CACCTCACACTTTCCTCAGGAGAGACACACCAACCATGACCTCAGCTGGGGACCCCAA 915

Db 18 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 32

Qy 916 CTCAGGCGCTCCAGCCCAACCTGCGCAGCCAGCCCTCGAAATGCAAGTCTTGTACGA 975

Db 32 nSerArgProSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrl 52

Qy 976 GTTTGAAGCTAGGAACCCACGGAACTGACTGTGTCTCCAGGAGAGAGCTGGAGTTCT 1035

Db 52 uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluVal 72

Qy 1036 GGACCACAGACGCGGTGGTGGTGAAGATCAGCGGGGACGCGGCTACATTCC 1095

Db 72 uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyrllePr 92

Qy 1096 AAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTGGGACCCAGGCCAGTCAACCTC 1155

Db 92 .oSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 112

Qy 1156 TCGGGTTCCAATGCTTCGACTTAGCTCGAGGCCCTGAAGAGGTTCACAGACTGGCTCAGGC 1215

Db 112 rArgValProMetLeuArgLeuSerSerArgProGluValThrAspTrpLeuGlnAl 132

Qy 1216 AGAGAACTTCTCCACTGCCAGCGGTGAGACACTTGGTCCCTGAGCGGGAGCCAGCTACT 1275

Db 132 aGluAsnPheSerThrAlaThrValrArgThrLeuGlySerLeuThrGlySerGlnLeuLe 152

Qy 1276 TCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCACCAATCTCTGC 1335

Db 152 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 172

Qy 1336 CCGGCTGGAGGCTGTTCAGAAGGATGCTGGGATAAGCCCT 1375

Db 172 rArgLeuGluAlaValArgMetLeuGlyIleSerPro 185

RESULT 6
AAU92275
ID AAM92275 standard; Protein; 348 AA.
XX

AC	AAM92275;		
XX		PR	14-SEP-2000; 2000US-0232400.
DT		PR	14-SEP-2000; 2000US-0232401.
XX	06-NOV-2001 (first entry)	PR	14-SEP-2000; 2000US-0233063.
XX		PR	14-SEP-2000; 2000US-0233064.
DE	Human digestive system antigen SEQ ID NO: 1624.	PR	14-SEP-2000; 2000US-0233065.
XX		PR	21-SEP-2000; 2000US-0234223.
KW	Human; digestive system antigen; gene therapy; cancer; appendicitis;	PR	21-SEP-2000; 2000US-0234274.
KW	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;	PR	25-SEP-2000; 2000US-0234997.
KW	digestive system disorder; Meckel's diverticulum.	PR	25-SEP-2000; 2000US-0234998.
XX		PR	26-SEP-2000; 2000US-0235484.
OS	Homo sapiens.	PR	27-SEP-2000; 2000US-0235834.
XX		PR	27-SEP-2000; 2000US-0235836.
PN	WO200155314-A2.	PR	29-SEP-2000; 2000US-0236327.
XX		PR	29-SEP-2000; 2000US-0236367.
PD		PR	29-SEP-2000; 2000US-0236368.
XX	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236369.
XX		PR	29-SEP-2000; 2000US-0236370.
PF	17-JAN-2001; 2001WO-US01324.	PR	02-OCT-2000; 2000US-0236802.
XX		PR	02-OCT-2000; 2000US-0237037.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0237038.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237039.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000; 2000US-0237039.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000; 2000US-0237040.
PR	16-MAR-2000; 2000US-0189874.	PR	13-OCT-2000; 2000US-0239935.
PR	17-MAR-2000; 2000US-0190076.	PR	13-OCT-2000; 2000US-0239937.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000; 2000US-0240960.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000; 2000US-0241221.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000; 2000US-0241785.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241786.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000; 2000US-0241808.
PR	07-JUL-2000; 2000US-0216880.	PR	20-OCT-2000; 2000US-0241809.
PR	11-JUL-2000; 2000US-0217487.	PR	20-OCT-2000; 2000US-0241826.
PR	11-JUL-2000; 2000US-0217496.	PR	01-NOV-2000; 2000US-0244617.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000; 2000US-0246474.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000; 2000US-0246475.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246610.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246611.
PR	18-AUG-2000; 2000US-0226279.	PR	08-NOV-2000; 2000US-0246613.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0226868.	PR	17-NOV-2000; 2000US-0249208.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249209.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249210.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249211.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000; 2000US-0249213.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000; 2000US-0249214.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249215.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000; 2000US-0249216.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000; 2000US-0249217.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249218.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249244.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249264.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249265.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000; 2000US-0249297.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000; 2000US-0249297.
PR	08-SEP-2000; 2000US-0232080.	PR	17-NOV-2000; 2000US-0249300.
PR	08-SEP-2000; 2000US-0232081.	PR	01-DEC-2000; 2000US-0250160.
PR	12-SEP-2000; 2000US-0231968.	PR	01-DEC-2000; 2000US-0250391.
PR	14-SEP-2000; 2000US-0232397.	PR	05-DEC-2000; 2000US-0251030.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000; 2000US-0251988.
PR	14-SEP-2000; 2000US-0232399.	PR	05-DEC-2000; 2000US-0256719.

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0246167.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
Rosen CA, Barash SC, Ruben SM;
WPI; 2001-581633/65.
N-PSDB; ABK43673.
XX
New isolated nucleic acid encoding a protein for diagnosing,
preventing, treating or ameliorating medical conditions and used as
food additives or preservatives -
XX
Claim 9; SEQ ID No 861; 837pp; English.
XX
The invention describes an isolated nucleic acid molecule (I) encoding a
novel central nervous system protein. (I) and polypeptides (II) encoded
by (I), are used to treat a medical conditions and in diagnosis of a
pathological condition. Disorders which are diagnosed or treated include
autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
angiogenesis, nervous system disorders e.g. Alzheimer's disease and
amyotrophic lateral sclerosis, infections caused by bacteria, viruses
e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
adenocarcinomas and irritable bowel syndrome, endocrine disorders e.g. diabetes
disorders e.g. testicular feminisation, disorders at the cellular level e.g.
leukaemia, disorders involving neovascularisation e.g. malignancies,
respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
acute kidney failure and blood related disorders e.g. myocardial
infarction. The polypeptides can also be used to aid wound healing and
epithelial cell proliferation, to prevent skin aging due to sunburn, to
maintain organs before transplantation, for supporting cell culture of
primary tissues, to regenerate tissues and in chemotaxis. The
polypeptides can also be used as a food additive or preservative to
increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:

Pred. No.: 3.42e-41 Length: 121
Score: 603.00 Matches: 116
Percent Similarity: 98.33% Conservative: 2
Best Local Similarity: 96.67% Mismatches: 2
Query Match: 19.28% Indels: 0
DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAU87343 (1-121)

Qy 1016 GGACGAGCTGGAGGTTCTGGACACAGCAGCGTGGTGGTGAAGAATGAGCG 1075
Db 2 GlyAspSerTrpGlnValLeuAlaHisSerLysArgTrpTrpValValValSAsnGluAla 21
Qy 1076 GGACGAGCGGCTACATTCCAAGCAACATCTCTGGAGCCCTTACAGCCGGGACCCCTGGG 1135

Db	22	GlyArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGly	14-AUG-2000; 2000US-0225270.
			PR 14-AUG-2000; 2000US-0225447.
QY	1136	ACCCAGGGCCAGTCACCCCTCGGGTTCCAAATGCTTCGACTTAGCTCGAGGCGCTGAAGAG	PR 14-AUG-2000; 2000US-0225757.
			PR 14-AUG-2000; 2000US-0225758.
Db	42	ThrInglyInSerProSerArgValProMetLeuArgLeuSerArgProGluGlu	PR 14-AUG-2000; 2000US-0225759.
			PR 18-AUG-2000; 2000US-0226279.
QY	1196	GTCACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCCGTGCAGGACACTGGGTCC	PR 22-AUG-2000; 2000US-0226681.
			PR 22-AUG-2000; 2000US-0226686.
Db	62	ValThrAspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySer	PR 22-AUG-2000; 2000US-0227182.
			PR 23-AUG-2000; 2000US-0227009.
QY	1256	CTGACGGGGAGCCAGCTACTTCGCATAGACCTGGGGAGCTACAGATGCTATGTCACAG	PR 30-AUG-2000; 2000US-0228924.
			PR 01-SEP-2000; 2000US-0229287.
Db	82	LeuThrGlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGln	PR 01-SEP-2000; 2000US-0229343.
			PR 01-SEP-2000; 2000US-0229344.
QY	1316	GAGGCCCAAGATCTGTCCCGCTGGAGGCTGTCCAGAGGATGCTGGGGATAAGCCCT	PR 01-SEP-2000; 2000US-0229345.
			PR 05-SEP-2000; 2000US-0229509.
Db	102	GluAlaProArgIleLeuSerArgLeuGluAlaValArgMetLeuGlyIleSerPro	PR 05-SEP-2000; 2000US-0229513.
			PR 06-SEP-2000; 2000US-0230437.
RESULT 8			PR 06-SEP-2000; 2000US-0230438.
AAU17191			PR 08-SEP-2000; 2000US-0231242.
ID AAU17191 standard; Protein; 121 AA.			PR 08-SEP-2000; 2000US-0231243.
XX AAU17191;			PR 08-SEP-2000; 2000US-0231244.
XX			PR 08-SEP-2000; 2000US-0231413.
XX			PR 08-SEP-2000; 2000US-0231414.
DT	07-NOV-2001 (first entry)		PR 08-SEP-2000; 2000US-0232080.
XX			PR 08-SEP-2000; 2000US-0232081.
DE			PR 12-SEP-2000; 2000US-0231968.
XX			PR 14-SEP-2000; 2000US-0232397.
XX			PR 14-SEP-2000; 2000US-0232398.
KW	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;		PR 14-SEP-2000; 2000US-0232399.
KW	antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;		PR 14-SEP-2000; 2000US-0232400.
KW	immune system disorder; rheumatoid arthritis; inflammatory condition;		PR 14-SEP-2000; 2000US-0232401.
KW	organ transplant rejection; infection; hepatitis C; blood disorder;		PR 14-SEP-2000; 2000US-0233063.
KW	sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;		PR 14-SEP-2000; 2000US-0233064.
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;		PR 14-SEP-2000; 2000US-0233065.
KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;		PR 21-SEP-2000; 2000US-0234223.
KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease;		PR 25-SEP-2000; 2000US-0234997.
KW	reproductive system; gastrointestinal; liver disorder; AIDS;		PR 25-SEP-2000; 2000US-0234998.
XX	acquired immune deficiency syndrome.		PR 26-SEP-2000; 2000US-0235484.
OS	Homo sapiens.		PR 27-SEP-2000; 2000US-0235834.
XX			PR 27-SEP-2000; 2000US-0235836.
PN	WO200154733-A1.		PR 29-SEP-2000; 2000US-0236327.
XX			PR 29-SEP-2000; 2000US-0236367.
PD			PR 29-SEP-2000; 2000US-0236368.
XX			PR 29-SEP-2000; 2000US-0236369.
PF	17-JAN-2001; 2001WO-US01312.		PR 29-SEP-2000; 2000US-0236370.
XX			PR 02-OCT-2000; 2000US-0236802.
XX	31-JAN-2000; 2000US-0179065.		PR 02-OCT-2000; 2000US-0237037.
PR	04-FEB-2000; 2000US-0180628.		PR 02-OCT-2000; 2000US-0237038.
PR	24-FEB-2000; 2000US-0184664.		PR 02-OCT-2000; 2000US-0237039.
PR	02-MAR-2000; 2000US-0186350.		PR 02-OCT-2000; 2000US-0237040.
PR	16-MAR-2000; 2000US-0198874.		PR 13-OCT-2000; 2000US-0239935.
PR	17-MAR-2000; 2000US-0190076.		PR 13-OCT-2000; 2000US-0239937.
PR	18-APR-2000; 2000US-0198123.		PR 20-OCT-2000; 2000US-0240960.
PR	19-MAY-2000; 2000US-0205515.		PR 20-OCT-2000; 2000US-0241221.
PR	07-JUN-2000; 2000US-0209467.		PR 20-OCT-2000; 2000US-0241785.
PR	28-JUN-2000; 2000US-0214886.		PR 20-OCT-2000; 2000US-0241786.
PR	30-JUN-2000; 2000US-0215135.		PR 20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216647.		PR 20-OCT-2000; 2000US-0241808.
PR	07-JUL-2000; 2000US-0216880.		PR 20-OCT-2000; 2000US-0241809.
PR	11-JUL-2000; 2000US-0217487.		PR 20-OCT-2000; 2000US-0241826.
PR	11-JUL-2000; 2000US-0217496.		PR 01-NOV-2000; 2000US-0244617.
PR	14-JUL-2000; 2000US-0218290.		PR 08-NOV-2000; 2000US-0246474.
PR	26-JUL-2000; 2000US-0220963.		PR 08-NOV-2000; 2000US-0246475.
PR	26-JUL-2000; 2000US-0220964.		PR 08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224518.		PR 08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0224519.		PR 08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225213.		PR 08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225214.		PR 08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225266.		PR 08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225267.		PR 08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225268.		PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254997.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465460/50.
XX N-PSDB; AAS27108.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
XX Claim 1; SEQ ID No 756; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction

CC pathway protein, amino acid sequences of the invention.
XX
XX
XX Alignment Scores: 4.14e-41 Length: 121
XX Pred. No.: 602.00 Matches: 116
XX Score: 98.33% Conservative: 2
XX Percent Similarity: 96.67% Mismatches: 2
XX Best Local Similarity: 19.25% Indels: 0
XX Query Match: 22 Gaps: 0
XX DB:
XX
XX US-09-762-021A-1 (1-1710) x AAU17191 (1-121)
QY 1016 GGAGAGAAGCTGGAGTCTTGACACAGCAAGCGTGGTGGCTGCAAGAATGAGCG 1075
DB 2 GlyAsp***trpGlnValLeuAspHisSerLysArgtrpIrpLeuValLysAsnGluAla 21
QY 1076 GGACGAGCGGCTACATTCCCAAGCAACATCTCGAGGCCCTTACAGCCGGGAGCCCTGGG 1135
DB 22 GlyArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGly 41
QY 1136 ACCCAGGGCCAGTCACCTCTCGGGTTCCTGCTTGCATCTAGCTTACGAGCCCTGAAGAG 1195
DB 42 ThrGlnGlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGlu 61
QY 1196 GTCACAGACTGGCTGCAGGAGAGAACTTCTCCACTGCACGGTGAGGACACTTGGGTCC 1255
DB 62 ValThrAspIrpLeuGlnAlaGluAsnPheserThrAlaThrValArgThrLeuGlySer 81
QY 1256 CTGACGGGGAGCCAGCTACTTCCGATAAGACCTGGGAGCTACAGATGCTATGTCCACAG 1315
DB 82 LeuThrGlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGln 101
QY 1316 GAGCCGCCAGATCTCTCCGGCTGGAGGCTGTCAGAGGATGCTGGGGATAGGCCCT 1375
DB 102 GluAlaProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 121
XX
XX RESULT 9
XX AAU87626
XX ID AAU87626 standard; Protein; 134 AA.
XX AC
XX AAU87626;
XX
XX 05-JUN-2002 (first entry)
XX
XX DE Novel central nervous system protein #536.
XX
XX Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
XX Homo sapiens.
OS
XX WO200155318-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01332.
XX
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.


```
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244675.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPT; 2001-465460/50.
DR N-PSDB; AAS27487.
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX Claim 1; SEQ ID No 1135; 880pp; English.
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Alignment Scores:
Pred. No.: 9,12e-41 Length: 134
Score: 598.00 Matches: 111
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 1
Query Match: 19.12% Indels: 0
DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAU17570 (1-134)
QY 827 CTTCCGGCGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAGACACAC 886
DB 14 LeuArgArgGlySerHisArgLeuGlySerThrSerHisPheProGlnGluysThrHis 33
QY 887 AACCATGACCCCTACGCTGGGACCCCAACTCCAGGCCCTCCAGCCCCCAAACTGCCAG 946
DB 34 AsnHisAspProGlnProGlyAspProAsnSerArgProSerSerProLysProAlaGln 53
QY 947 CCAGCCCTGAANAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACCGGAACGTACT 1006
DB 54 ProAlaLeuLysMetGlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThr 73
QY 1007 GTGCTCCAGGGAGAGAAGCTGGAGGTTCTTGACACACAGACGCGTGGTGGCTGGAAG 1066
DB 74 ValValGlnGlyGluLysLeuGluValLeu***HisSerLysArgTrpTrpLeuValLys 93
QY 1067 AATAGCGGGGACGAGCGGCTACATTCACAGACATCTCTGGAGCCCTACAGCGGGG 1126
DB 94 AsnGluAlaGlyArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGly 113
QY 1127 ACCCGTGGGACCCAGGCGCAGTCACCCCTCTCGGTT 1162
DB 114 ThrProGlyThrGlnGlyGlnSerProSerArgVal 125
```

RESULT 11

AAAM25595
ID AAAM25595 standard; Protein; 535 AA.
XX AC AAAM25595;
XX DT 16-OCT-2001 (first entry)
XX DE Human protein sequence SEQ ID NO:1110.
XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; virucide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiagregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; antiparkinsonian; infection;
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.
XX OS Homo sapiens.
XX PN WO200153455-A2.
XX PD 26-JUL-2001.
XX PF 22-DEC-2000; 2000WO-US35017.
XX PR 23-DEC-1999; 99US-0471275.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WP1; 2001-457603/49.
XX DR N-PSDB; AAH99536.
XX PT Isolated human polynucleotides encoding polypeptides, useful for the
PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -
XX PS Claim 20; Page 228; 1217pp; English.
XX CC AAH99166 to AAH99904 encode the human proteins given in AAAM25225 to
CC AAAM25963. The proteins can have activities based on the tissues and
CC cells they are expressed in, such as: antinflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerary;
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX CC neurological disorders.
XX Sequence 535 AA;

Alignment Scores:				
Pred. No.:	2.03e-39	Length:	535	
Score:	584.00	Matches:	163	
Percent Similarity:	45.00%	Conservative:	71	
Best Local Similarity:	31.35%	Mismatches:	156	
Query Match:	18.68%	Indels:	130	
DB:	22	Gaps:	19	
US-09-762-021A-1 (1-1710) x AAM25595 (1-535)				
Qy	131	GAGCAGGCACGCTATCTGGAGCGGGATCCCTCCAGACAGACCCACAGAGGCCCTA	190	
Db	10	GlulyslleArgGlnArgGlnSerIleLeuProPro	21	
Qy	191	GAGCAGACGCTCCACCATCCCAAGGCCCTGCCA---CGCCACACACAGTCCCGGAA	247	
Db	22	-----ProGlnGlyProAlaProIleProPheGlnHisArgGlyGlyAspSer	37	
Qy	248	CAAAGTGCC-----TTTACTCTGCCTCTCTCCAAAGCGG	280	
Db	38	ProGluAlaLysAsnArgValGlyProGlnValProLeuSerGluProGlyPheArgArg	57	
Qy	281	TCCTCTTCCCGCCGAGGCCA-----GAGAGGACGAGGAA	316	
Db	58	ArgGluSerGlnGluProArgAlaValLeuAlaGlnLysIleGluLysGluThrGln	77	
Qy	317	GTGCTGAACCATGTCCTAAGGGACATTCAGCTGTTCATGGGAAGCTGGAGAAGGCC---	373	
Db	78	IleLeuAsnCysAlaLeuAspIleGluTrpPheValAlaArgLeuGlnLysAlaAla	97	
Qy	374	-----CAGGCAAGACACAGCAGGAAGAANAATTTGGGAAAAA	412	
Db	98	GluAlaPheLysGlnLeuAsnGlnArgLysLysLysLysLys-----GlyLysLys	115	
Qy	413	AACAAGGACACAGGAGGTCTCAC-----CAGGCACAGTACATT	451	
Db	116	AlaProAlaGluGlyValLeuThrLeuArgAlaArgProProSerGluGlyGluPheIle	135	
Qy	452	GACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAAGCTGGCCACCTGGCTG	511	
Db	136	AspCysPheGlnLysIleLysLeuAlaIleAsnLeuLeuAlaLysLeuGlnLysHisIle	155	
Qy	512	AAGCAGACAGTGCCTGAGCTGTACACATCTCTTCAAGTCCCTCAAGTTCATCTCTG	571	
Db	156	GlnAsnProSerAlaAlaGluLeuValHisPheLeuPheGlyProLeuAspLeuIleVal	175	
Qy	572	GCCAGGTCCCTGAGGTGGCTAGCAGCCCAAGTGATCTCACCCCTCTCTCACCCCTAAA	631	
Db	176	AsnThrCysSerGlyProAspIleAlaArgSerValSerCysProLeuLeuSerArgAsp	195	
Qy	632	GCTATCAACCTGTACAGTCTGTCTAAGCCCACTGAGAGTAACTTTGGATGGGTGTG	691	
Db	196	AlaValAspPheLeuArgGlyHisLeuValProLysGluMetSerLeuTrpGluSerLeu	215	
Qy	692	GGCCAGCCTGGACCACTAGCCGGGCGACGTGACAGCGCATGAGCCCTGCC---TAC	748	
Db	216	GlyGluSerTrpMetArgProArgSerGluTrpProArgGluProGlnValProLeuTyr	235	
Qy	749	CAACCCACATCTCAGATGACTGGCAACTTCCA-----	781	
Db	236	ValProLysPheHisSerGlyTrpGluProProValAspValLeuGlnAlaProTrp	255	
Qy	782	-----GAGCCCTCCAGCCAGCACCTTAGGATACAGGACCTGTTCCCTTCGGCGG	835	
Db	256	GluValGluGlyLeuAlaSerAlaProIleGluValSerProValSer-----Arg	273	
Qy	836	GGAAGTCATAGTTAGGAGCACCTCACACTTCTCCTCAGGAGAGACACACACCATGAC	895	
Db	274	GlnSerIleArg---AsnSerGlnLysHisSerPro-----ThrSerGluProThr	289	
Qy	896	CCTCAGCTGGGGAC-----CCCAACTCCAGGCCCTCCAGCCCAACCTGCCAG	946	

Db 336 AsnIleAlaAspProSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet 355
 QY 566 ATCTGCGCCAGGTGCGCTGAGGTGCGCTAGCAGGCCAAGTATCTCACCCCTCTCCACC 625
 Db 356 IleValAsnThrSerGlyGlyProGluPheAlaSerSerValArgArgProHisLeuThr 375
 QY 626 CCTAAAGCTATCAACTGTCTACAGTCTCTTAAGCCACCTGAGATACCTTTGGATG 685
 Db 376 SerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsnGluLeuTrpThr 395
 QY 686 GGGTTGGCCCGCCCTGGACCACTAGCCGGCGGCGACTGCAGCGCGATGAGCCCTGCC 745
 Db 396 SerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGluGlyProPro 415
 QY 746 TACCAACCCACATCTTCAGATGCTGCACTGCACTTCCA-----GAGCCCTCCAGCAAGCA 799
 Db 416 TyrArgProGluPhePheSerGlyTrpGluProProValThrAspProGlnSerArgAla 435
 QY 800 CCTTATAGGATACAGACCTCTTCCCTCGCGGGGGAAGTCATAGTTAGGGAGCAC 859
 Db 436 -----TrpGluAspProValGluLysGlnLeuGlnHisGluArgArgArgGln 452
 QY 860 TCACACTTTCTCAGGAGACACACACACCAT---GACCCTCAGCTGGGGACCCCAAC 916
 Db 453 GlnSerAlaProGlnValAlaValAsnGlyHisArgAspLeuGluPro----- 468
 QY 917 TCCAGGCCCTCCAGGCCCAACCTGCGCCAGCCGCTGAAATGCAAGCTTG----- 970
 Db 469 -----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn 486
 QY 971 TAGGAGTTTGAAGCTAGGAAACCCACCGGAAGTCTGTGTCAGGGAGAGAGCTGGAG 1030
 Db 487 TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu 506
 QY 1031 GTTCTGCACACAGACGGTGGTGGTGGTGAAGATGAGCGCGGAGCGGCGCTAC 1090
 Db 507 ValLeuAspSerArgLysTrpTrpLysValArgAspProAlaGlyGlnGluGlyTyr 526
 QY 1091 ATTCCAAGCAACATCTGGAGCCCTACAGCGGGG----- 1126
 Db 527 ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisSerGlnSer 545
 QY 1127 -----ACCCCTGGG----- 1135
 Db 546 ProAlaArgSerLeuAsnSerThrProProProProAlaProAlaProPro 565
 QY 1135 ----- 1135
 Db 566 ProAlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerLeuAsn 585
 QY 1135 ----- 1135
 Db 586 GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluLeu 605
 QY 1136 -----ACCGAGGCCAGTCA---CCCTCTCGGGTT----- 1162
 Db 606 GlnAlaArgLeuAlaGlnArgSerGlyProSerArgAlaValProGlyProArgAla 625
 QY 1163 -----CCAATGCTTCCAGCTAGCTCGAGCGCTGAGAGGTCACAGCTCGGTCAGGCA 1216
 Db 626 ProGluProGlnLeuSerProGlySerAlaSerGluValArgAlaTrpLeuGlnAla 645
 QY 1217 GAGAATCTTCCACTGCCAGGTGAGACACTTGGCTCCCTGACGGGGAGCCAGCTACTT 1276
 Db 646 LysGlyPheSerSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnLeuPhe 665
 QY 1277 CGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACGAATCTGTCC 1336
 Db 666 SerLeuGlnArgGluLeuLeuArgAlaValSerProGluGluGlyAlaArgValTyrSer 685
 QY 1337 CGGCTGGAGCTCTCAGAGGATGCTGGG 1366
 Db 686 GlnValThrValGlnArgSerLeuLeuGly 695

RESULT 13

AAU31985
 ID AAU31985 standard; Protein; 569 AA.

XX AAU31985;

XX 18-DEC-2001 (first entry)

DE Novel human secreted protein #2476.

XX Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

PN WO200179449-A2.

XX 25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.

PR 18-APR-2000; 2000US-0552929.

PR 26-JAN-2001; 2001US-0770160.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-611725/70.

Nucleic acids encoding a range of human polypeptides, useful in genetic vaccination, testing and therapy -

Claim 20; Page 535-536; 765pp; English.

The invention relates to novel human secreted polypeptides. The polypeptides and antibodies to the polypeptides are useful for determining the presence of or predisposition to a disease associated with altered levels of polypeptide. The polypeptides are also useful for identifying agents (agonists and antagonists) that bind to them. Cells expressing the proteins are useful for identifying a therapeutic agent for use in treatment of a pathology related to aberrant expression or physiological interactions of the polypeptide. Vectors comprising the nucleic acids encoding the polypeptides and cells genetically engineered to express them are also useful for producing the proteins. The proteins are useful in genetic vaccination, testing and therapy, and can be used as nutritional supplements. They may be used to increase stem cell proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon and/or nerve tissue growth or regeneration; immune suppression and/or stimulation; as anti-inflammatory agents; and in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid sequences of novel human secreted proteins of the invention.

Sequence 569 AA;

Alignment Scores:
 Pred. No.: 1,61e-37 Length: 569
 Score: 561.00 Matches: 164
 Percent Similarity: 43.46% Conservative: 82
 Best Local Similarity: 28.98% Mismatches: 188
 Query Match: 17.94% Indels: 133
 DB: 22 Gaps: 16

US-09-762-021a-1 (1-1710) x AAU31985 (1-569)

QY 131 GAGCAGGCACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCACCAGAGACCCTA 190

Db 10 GluLysIleArgGlnArgGlnSerIleLeuProPro----- 21

QY 191 GAGCAGAGCTCCCAACCATCCCAAGGCCCTGCCACGCCACACAGTGC----- 241
 ||||| ||| ||||| |||||

Db 22 -----ProGlnGlyProAlaProIleProSerSerThrAlaAlaGlyIlePro 37
Qy 242 CGAAGACCAAGT-----GCCTTTACTCTGCCTCCCTCCCAAGCGG 280
Db 38 ArgArgProArgIleAlaTrpAlaArgArgTrpProLeuSerGluProGlyPheArgarg 57
Qy 281 TCCTCTTCCCGAGGACCCA-----GAGAGGGAGGAGGAA 316
Db 58 ArgGluSerGlnGluProArgAlaValLeuAlaGlnLysIleGluLysGluThrGln 77
Qy 317 GTGCTGAACCATGTCTAAGGACATTCAGCTGTTTCATGGGAAGCTGGGAAGGCC--- 373
Db 78 IleLeuAsnCysAlaLeuAspIleGluTrpPheValAlaAlaArgLeuGlnLysAlaAla 97
Qy 374 -----CAGCAAGACACAGCAGGAAGAGAATTTGGGAAGAAA 412
Db 98 GluAlaPheLysGlnLeuAsnGlnArgLysLysGlyLysLysLys-----GlyLysLys 115
Qy 413 AACAAAGGACAGGAGGTCTCACC-----CAGGCACAGTACATT 451
Db 116 AlaProAlaGluGlyValLeuThrLeuArgAlaArgProProGluGlyGluPheIle 135
Qy 452 GACTCTCTCCAGAGATCAAGTACAGCTTCAACCTCTCGGGAAGCTGGCCACCTGGCTG 511
Db 136 AspCysPheGlnLysIleLysLeuAlaIleAsnLeuLeuAlaLysLeuGlnLysHisIle 155
Qy 512 AAGGAGACAAGTGCCTGAGCTGCTACATCTCTTCAAGTCCCTGAACTTCATCTG 571
Db 156 GlnAsnProGlnArgArgAspValValHisPheLeuPheGlyProLeuAspLeuIleVal 175
Qy 572 GCCAGGTGCCCTGAGCGTGGCTGACAGCCCAAGTGATCTCACCCCTCTCCACCCCTAAA 631
Db 176 AsnThrCysSerGlyProAspIleAlaArgSerValSerCysProLeuPheSerArgAsp 195
Qy 632 GCTATCAACCTGTACAGTCTCTGTCTAGCCACCTGAGATACCTTTGATGGGGTTG 691
Db 196 AlaValAspPheLeuArgGlyHisLeuValProLysGluMetSerLeuTrpGluSerLeu 215
Qy 692 GSCCCAGCTCGACCACTAGCCGGCCGACCTGGACAGCGGATGACCCCTGCC---TAC 748
Db 216 GlyGluSerTrpMetArgProArgSerGluTrpProArgGluProGlnValProLeuTrp 235
Qy 749 CAACCCACATTCTCAGATGACTGGCACTTCCA----- 781
Db 236 ValProLysPheHisSerGlyTrpGluProProValAspValLeuGlnLysAlaProTrp 255
Qy 782 -----GAGCCCTCCAGCAAGCACCTTAGATACAGGACCCCTGTTCCCTTCGGCGG 835
Db 256 GluValGluGlyLeuAlaSerAlaProIleGluGluValSerProValSer-----Arg 273
Qy 836 GGAAGTCATAGGTTAGGAGCACCTCACACTTTCCTCAGGAGAGACACACACCATGAC 895
Db 274 GlnSerIleArg---AsnSerGlnLysHisArgProHisPheArgAlaHis----- 289
Qy 896 CCTCAGCTGGGGACCCCAAC-----TCCAGCCCTCCACAGC 931
Db 290 ProProGlyGlyCysProTyHisGlnSerAlaProHisIleLeuThrArgGlyTrpGln 309
Qy 932 CCCAAACCTGCCCGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAAC 991
Db 310 ProThrProAlaMet---AlaLysTrpValLysIleLeuTyAspPheThrAlaArgAsn 328
Qy 992 CCACGGGAACCTGCTGTGCTCCAGGAGAGAAAGCTGGAGGTTCTGGACCACAGCAAGCGG 1051
Db 329 AlaAsnGluLeuSerValLeuLysAspGluValLeuGluValLeuAspGlyArgGln 348
Qy 1052 TGGTGGCTGGTGAAGATCAGCGGCGGAGCGGCTACATTCCAAGCAACATCTGGAG 1111
Db 349 TrpTrpLysLeuArgSerArgSerGlyGlnAlaGlyTrpValProGlyAsnIleLeuGly 368
Qy 1112 CCCCTACAGCGC----- 1126
Db 369 GluAlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGlyGlnLysTrpLeuGly 388

Qy 1127 ACCCTGGAGCCAGGCGCAGTCACCTCTCGGGTTCCA----- 1165
Db 389 ThrProGlnProAspProGlnAlaThrProLysLeuProGlyGlyThrIleAspGluLeu 408
Qy 1165 ----- 1165
Db 409 MetGlnHisMetAspGluValAsnAspGluLeuIleArgLysIleThrThrSerArgAla 428
Qy 1166 -----ATCCTTCGACTT 1177
Db 429 GlnProGlnArgHisPheArgValGluArgSerGlnProValSerGlnProLeuThrTrp 448
Qy 1178 AGCTCGAGGCTGAAGAGGTCCACAGCTGGCTGAGGAGAGAACTTCTCCACTGCCACG 1237
Db 449 GluSerGlyProAspGluValArgAlaPheLeuGluAlaLysAlaPheSerProArgIle 468
Qy 1238 GTGAGGACACTTGGTCCCTGACGGGAGCAGCTACTTCCCATTAAGACCTGGGGAGCTA 1297
Db 469 ValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLysGluGluLeu 488
Qy 1298 CAGATGCTATGTCACAGGAGGCCCCAGCAATCTGTCGGCTGGAGGCTGTGAGAAGG 1357
Db 489 LysLysValCysGlyGluGluGlyPheArgValTrpSerGlnLeuThrMetGlnLysAla 508
Qy 1358 ATGCTGGG-GATAAGCCCTTAGCCAGCAGCTTAGACACCTCCAAGAACCCGCGCTG 1416
Db 509 PheLeuGluLysGlnGlnSerGlySerGluLeuGluGluLeuMetAsnLysPheHisSer 528
Qy 1417 ATGCAAGATGCCAGATCTGATACCCATTTAGAGCCCCGAGAAATTCCTCTTCTGGATCCCAG 1476
Db 529 MetAsnGlnArgArgGlyGluAspGlnLeuGlyProAlaAlaLeuGlyTrpGlyLeuArg 548
Qy 1477 TTTCAGCAACCCACACA 1494
Db 549 ArgGlySerProProThr 554
RESULT 14
ID AAR35451 standard; Protein: 821 AA.
AC AAR35451;
XX XX
DT 17-DEC-2001 (updated)
DT 25-AUG-1993 (first entry)
XX XX
DE Mouse eps8.
XX XX
KW Epidermal growth factor receptor; EGFR-pathway substrate; eps;
KW tyrosine kinase receptor; TKR; SH2; SH3; mitogenesis.
XX XX
OS Mus musculus.
XX XX
PN USN7935311-N.
PD 01-APR-1993.
XX XX
PF 25-AUG-1992; 92US-0935311.
XX XX
PR 25-AUG-1992; 92US-0935311.
XX XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX XX
PI Di Fiore PP, Fazloli F;
XX XX
DR WPI; 1993-159477/19.
XX XX
DR N-PSDB; AAQ40730.
XX XX
PT Epidermal growth factor receptor substrate, eps 8 - used to
enhance mitogenic response of cells to epidermal growth factor
XX XX
PS Disclosure; Page 30-37; 40pp; English.
XX XX

CC Eps8 is a novel EGFR substrate. The protein bears the
 CC characteristic signatures of TKR substrates including SH2 and
 CC SH3 domains. Eps8 is involved in the transduction of mitogenic
 CC signals and it can be used to enhance the mitogenic response of
 CC cells to EGF.
 CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US
 CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
 CC
 XX
 SQ Sequence 821 AA;

Alignment Scores:

Pred. No.:	2,91e-37	Length:	821
Score:	558.50	Matches:	158
Percent Similarity:	38.82%	Conservative:	78
Best Local Similarity:	25.99%	Mismatches:	171
Query Match:	17.86%	Indels:	201
DB:	14	Gaps:	17

US-09-762-021A-1 (1-1710) x AAR35451 (1-821)

```

Qy 101 GGGCTGCTATGGAAGCCCTCCCTATGGAGCAGGCACCTATCTGGAGCCGGGATC 160
Db 190 GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209
Qy 161 CTTCCAGAACGCCCCACGAGGAGCCCTAGAGCAGCAGCCCTCCACCATCCCAAGGCC 220
Db 210 ProProPro-----ProArgAlaProAlaPro 218
Qy 221 CTGCCAGCCACACC-----AGTGCC 241
Db 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTyrSerAla 238
Qy 242 CGAAGAACCAAGTCTTACTCTCCCTCCCAAGGGGCTCTCTCCCGGAGAC---- 298
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
Qy 299 CCAGAG-----AGGACGAGGAAGTGTGAACCATGTCTCTAAGG 337
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
Qy 338 GACATTGAGCTGTTTCATGGGAAGCTGGAGAGCCCGCAG-----GCA 379
Db 279 AspIleGluPheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
Qy 380 AAGACCAGCAGAGAGAGAAATTTGGGAAAAAACAAGGACGAGGAGGTCTCACC---- 436
Db 299 LysArgLysLysSerLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
Qy 437 -----CAGGCACAGTACATTGACTGCTTCCAGAAAGATCAAGTACAGC 478
Db 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
Qy 479 TTCACCTCTCTGGGAAGCTGGCCACCTGGCTGAAGAGAGACAAGTCCCTGAGCTGCTGA 538
Db 339 PheAsnLeuLeuAlaLysLeuLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
Qy 539 CACATCTCTTCAAGTCCCTCAACTTCTCTGCGCAGGTCCTGAGGCTGGCCCTAGCA 598
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
Qy 599 GCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTA 658
Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
Qy 659 AGCCACCTGAGTAACCTTTGGATGGGTTGGGCCCGCAGCTGACCACTAGCCGGGCC 718
Db 399 ThrAlaGluGluArgLysLeuTyrMetSerLeuGlyAspSerTrpValLysValArgAla 418
Qy 719 GACTGGACAGCGGATGAGCCCTG---CCCTACCAACCCACATTCTCAGATGAGTGGCAA 775
Db 419 GluTrpProLysGluGlnPheIleProProTyrValProArgPheArgAsnGlyTrpGlu 438

```

```

Qy 776 CTTCCA-----GAGCCCTCCAGCAAGCACCTTAGGTATACGAGC 817
Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
Qy 818 CCTGTTTCCCTTCGCGGGGAAGTCAT---AGTTAGGAGCAGCCTCACACTTCTCTCAG 874
Db 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
Qy 875 GAGAAGACACACACCATGACCCCTCAGCTGGGAC----- 910
Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerSerMet 496
Qy 911 -----CCC 913
Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
Qy 914 AAC-----TCCAGGCGCTCCAGCCGCCAACACTGCCAGCAGCCCTGAAATGCAA 964
Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
Qy 965 GTCCTGTACAGTTTGAAGCTAGGAACCCAGCGGAAGTCTGCTGCCAGGGAGAGAG 1024
Db 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal 556
Qy 1025 CTGGAGGTTCTGGACACAGCAAGCGTGTGCTGTGAAGATGAGCGCGGAGCGGAGC 1084
Db 557 LeuGluLeuAspArgArgGlnTrpLysValArgAsnAlaSerGlyAspSer 576
Qy 1085 GGCTACATTCCAAGCAACATCTCGAG----- 1111
Db 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
Qy 1111 ----- 1111
Db 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
Qy 1112 -----CCCTCA 1117
Db 617 AlaAspThrProSerAlaProSerProProThrProAlaProValProValProLeu 636
Qy 1118 CAGCGGGGACCCCTGGG----- 1135
Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
Qy 1135 ----- 1135
Db 657 AsnSerSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
Qy 1135 ----- 1135
Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
Qy 1136 ---ACCCAGGGCCAGTCACCTCTCGG-----GTTCCA 1165
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
Qy 1166 ATGCTTCGACTTAGC-----TCGAGGCGCTGAAGAGGTTCACAGACTGGCTGCAGCGCAG 1219
Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
Qy 1220 AACTTCTCCACTGCCAGCGGTGAGCAGACTTGGGTCTCTGCGGAGCGAGCAGCTACTTCGC 1279
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
Qy 1280 ATAAGACCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCGCCAGCAATCTCTCCCGG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValCysPro---GluGlyAlaArgValPheAsnGln 775
Qy 1340 CTGAGGCTGTCAAGAGGATGCTG 1363
Db 776 IleThrValGlnLysAlaAlaLeu 783

```

RESULT 15
AAG75249
ID AAG75249 standard; Protein; 112 AA.
XX
AC AAG75249;
XX
DT 03-SEP-2001 (first entry)
XX
Human colon cancer antigen protein SEQ ID NO:6013.
XX
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR N-PSDB; AAH34654.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7506-7508; 9803pp; English.
XX
AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
cancer-associated nucleic acid molecules (N) and proteins (P), where
the proteins are collectively known as colon cancer antigens. The colon
cancer antigens have cytostatic activity and can be used in gene
therapy and vaccine production. N and P may be used in the prevention,
diagnosis and treatment of diseases associated with inappropriate P
expression. For example, N and P may be used to treat disorders
associated with decreased expression by rectifying mutations or deletions
in a patient's genome that affect the activity of P by expressing
inactive proteins or to supplement the patients own production of P.
Additionally, N may be used to produce the colon cancer-associated Ps,
by inserting the nucleic acids into a host cell and culturing the cell
to express the proteins. N and P can be used in the prevention, diagnosis
and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
and AAB77789 represent sequences used in the exemplification of the
present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 112 AA;

Alignment Scores:
Pred. No.: 2,98e-31 Length: 112
Score: 482.00 Matches: 94
Percent Similarity: 96.00% Conservative: 2
Best Local Similarity: 94.00% Mismatches: 4
Query Match: 15.41% Indels: 0
DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAG75249 (1-112)

QY 311 GAGGAAGTCTGAACCATTCCTCAAGGACATTGAGCTGCTTCAAGGAAAGCTGGAGAAG 370
DB 13 GInGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetCglyLysLeuGluLys 32
QY 371 GCCCAGGCAAGACCAGCAGGAAGAAGAAATTTGGGAAAAAACAAGGACCAGGGAGGT 430
* * *

Db 33 AlaGlnAlaLysThrSer**LysLysLysPheGlyLysAsnLysAspGlnGlyGly 52
QY 431 CTCACCCAGGCACAGTACATTGACTGCTTCCAGAAAGATCAAGTACAGCTTCAACCTCCTG 490
Db 53 LeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysHisSerPheAsnLeuLeu 72
QY 491 GGAAGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCTGCTACACATCCTCTTC 550
Db 73 GlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHisIleLeuPhe 92
QY 551 AAGTCCCTGAACCTTCATCTCTGGCCAGGTGCCCTGAGGCTGGCTAGCAGCCCAAGTGATC 610
Db 93 Lys***LeuAsnPhe***LeuAlaArgCysProGluAlaGly**AlaAlaGlnValIle 112

Search completed: February 25, 2003, 19:39:48
Job time : 53 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 17:06:18 ; Search time 17.5 Seconds

(without alignments)
5750.080 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 3127

Sequence: 1 ggcagagcactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09762021/runat_24022003_153039_7605/app_query.fasta_1.1863
-DB-Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021.ecgn1.1.9.0runat_24022003_153039_7605 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	558.5	17.9	821	1	US-07-935-311A-4
2	558.5	17.9	821	1	US-08-368-079-4
3	558.5	17.9	821	5	PCT-US93-07996-4
4	200.5	6.4	1057	3	US-08-931-820-4
5	191	6.1	1078	3	US-08-963-825-21
6	191	6.1	1078	4	US-09-500-811-21
7	191	6.1	1078	4	US-09-570-573-21
8	191	6.1	1078	4	US-09-548-608-21
9	181.5	5.8	561	1	US-08-842-255-52
10	179.5	5.7	504	4	US-09-219-849-3
11	178	5.7	552	4	US-09-219-849-7
12	177.5	5.7	1461	4	US-09-585-887-9

13	177.5	5.7	1461	4	US-09-289-578-9	Sequence 9, Appli
14	176.5	5.6	633	1	US-08-642-255-73	Sequence 73, Appli
15	176.5	5.6	720	4	US-09-219-849-4	Sequence 4, Appli
16	176.5	5.6	777	1	US-08-642-255-53	Sequence 53, Appli
17	176.5	5.6	960	4	US-09-219-849-5	Sequence 5, Appli
18	175.5	5.6	1341	3	US-08-963-825-18	Sequence 18, Appli
19	175.5	5.6	1341	4	US-09-500-811-18	Sequence 18, Appli
20	175.5	5.6	1341	4	US-09-570-573-18	Sequence 18, Appli
21	175.5	5.6	1341	4	US-09-548-608-18	Sequence 18, Appli
22	174.5	5.6	1057	3	US-08-931-820-1	Sequence 1, Appli
23	172.5	5.5	1185	4	US-09-041-886-23	Sequence 23, Appli
24	171.5	5.5	960	4	US-09-219-849-6	Sequence 6, Appli
25	171.5	5.5	1274	4	US-09-095-443-2	Sequence 2, Appli
26	169	5.5	720	4	US-09-219-849-4	Sequence 4, Appli
27	169	5.5	777	1	US-08-642-255-53	Sequence 53, Appli
28	168.5	5.4	1065	1	US-08-642-255-72	Sequence 72, Appli
29	167	5.4	504	4	US-09-219-849-3	Sequence 3, Appli
30	167	5.4	561	1	US-08-642-255-52	Sequence 52, Appli
31	166	5.3	1060	3	US-08-931-820-3	Sequence 3, Appli
32	166	5.3	1418	3	US-08-963-825-20	Sequence 20, Appli
33	166	5.3	1418	4	US-09-010-999-1	Sequence 1, Appli
34	166	5.3	1418	4	US-09-500-811-20	Sequence 20, Appli
35	166	5.3	1418	4	US-09-570-573-20	Sequence 20, Appli
36	163.5	5.2	1442	2	US-09-548-608-20	Sequence 12, Appli
37	163.5	5.2	1442	2	US-08-316-650-12	Sequence 12, Appli
38	163.5	5.2	1442	5	PCT-US95-02251-12	Sequence 12, Appli
39	162	5.2	822	4	US-09-219-849-49	Sequence 49, Appli
40	158	5.1	1054	1	US-08-642-255-62	Sequence 62, Appli
41	157	5.0	357	1	US-07-609-716-66	Sequence 66, Appli
42	157	5.0	357	1	US-08-642-255-33	Sequence 33, Appli
43	157	5.0	357	4	US-08-475-411A-66	Sequence 66, Appli
44	157	5.0	357	4	US-08-478-029A-66	Sequence 66, Appli
45	155.5	5.0	1418	4	US-09-010-999-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-07-935-311A-4

; Sequence 4, Application US/07935311A
; Patent No. 5378809
; GENERAL INFORMATION:
; APPLICANT: Di Fiore, Pier Paolo
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: Substrate of the Epidermal Growth
; TITLE OF INVENTION: Factor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,311A
; FILING DATE: 19920825
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH035.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 821 amino acids

; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-935-311A-4

Alignment Scores:

Pred. No.: 4,16e-39 Length: 821
Score: 558.50 Matches: 158
Percent Similarity: 38.82% Conservative: 78
Best Local Similarity: 25.99% Mismatches: 171
Query Match: 17.86% Indels: 201
DB: 1 Gaps: 17

US-09-762-021A-1 (1-1710) x US-07-935-311A-4 (1-821)

```
QY 101 GGCCTGTATGAAAGCGCGTCCCTATGGAGCAGCGCATCTATCGAGCGGGGATC 160
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 161 CCTCCAGAACAGCCGCCACAGAGGACCTTAGAGCACAGCCTCCACCATCCCAAGGCC 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 ProProPro-----ProArgAlaProAlaPro 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 221 CTGCCAGCCACACC-----AGTGC 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaAlaTrpSerAla 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 242 CGAGAACCAAGTGCCTTTACTCTGCTCTCCAGCGCGTCTCTTCCCGCAGGAC--- 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 299 CGAGAG-----AGGGACGAGGAAGTGTCTGAACCATCTCCTAAAGG 337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuasp 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 GACATTGAGCTTTCATGGGAAGCTGGAGTGGAGAGGCCAG-----GCA 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 AspIleGluPhePheIleThrLysLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 380 AAGACCAGCAGCAGAGAAATTTGGGAAAAAACAAGGACCGGAGGCTCTCACC--- 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299 LysArgLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 437 -----CAGCACAGTACATTGACTGCTCCAGAAAGTCAAGTACAGC 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 ArgAlaLysProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 TTCACCTCTCGGAGGCTGCCACCTGGCTGAGAGGACAAAGTGCCTCAGCTCGTA 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 PheAsnLeuAlaLysLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 539 CACATCTCTTCAAGTCCCTGAACCTTCATCTCTGCGGAGGTGCGCTGAGGCTGCGCTAGCA 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 599 GCCCAAGTGATCTACCCCTCTCACCCTTAAAGCTATCAACCTGCTACAGTCCCTGCTCTA 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 659 AGCCACCTGAGAGTAACCTTTGGATGGGTGGGCCAGCCCTGGACCACTAGCCGGGCC 718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 719 GACTGACAGCGGATGAGCCCTG---CCCTACCAACCCACATTTCTCAGATGACGTGCA 775
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 GluTrpProLysGlnPheIleProProTyrValProArgPheArgAsnGlyTrpGlu 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 776 CTTCAC-----GAGCCCTCCAGCCAGCACCTTAGATACCAAGGAC 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 ProPrometLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 CCTGTTTCCCTTCGGGGGAAGTCAAT---AGTTAGGAGGACACCTCACACTTCCCTCAG 874
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 2

US-08-368-079-4
; Sequence 4, Application US/08368079
; Patent No. 5610018
; GENERAL INFORMATION:
; APPLICANT: Di Fiore, Pler Paolo
; APPLICANT: Fazioli, Francesca

TITLE OF INVENTION: eps8, A Substrate for the Epidermal Growth Factor
TITLE OF INVENTION: Kinase, Antibodies Thereto, and Methods of Use Thereof

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear

STREET: 620 Newport Center Drive, 16th Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/368,079

FILING DATE: 03-JAN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/935,311

FILING DATE: 25-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Israel, Ned A.

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH035.001DV1

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 821 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-368-079-4

Alignment Scores:

Pred. No.:	4.16e-39	Length:	821
Score:	558.50	Matches:	158
Percent Similarity:	38.82%	Conservative:	78
Best Local Similarity:	25.99%	Mismatches:	171
Query Match:	17.86%	Indels:	201
DB:	1	Gaps:	17

US-09-762-021a-1 (1-1710) x US-08-368-079-4 (1-821)

Qy	101	GGGCTGCTATGGAAGCCGCTCCCTATGGAGCAGGACGCTATCTGGAGCCGGGATC	160
Db	190	GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle	209
Qy	161	CTCCAGAACACCCACAGAGGACCTTAGACACACCTCCACCATCCCCAGGCCC	220
Db	210	ProProPro-----ProArgAlaProAlaPro	218
Qy	221	CTGCCACCCACAC-----AGTGCC	241
Db	219	ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla	238
Qy	242	CGAGAACCAAGTGCTTACTCTGCTCTCCCAAGCGGCTCTCTTCCCGGAGGAC---	298
Db	239	TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluLuphr	258
Qy	299	CCAGAG-----AGGACGAGGAGGAGTCTGACCATGCTCCTTAGG	337
Db	259	ProGluMetMetAlaAlaArgIleAspArgValGlnIleLeuAsnHisIleLeuAsp	278
Qy	338	GACATTGAGCTTCTCGGAAAGCTGGAGAGGCCAG-----GCA	379
Db	279	AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer	298
Qy	380	AAGACACAGGAGGAAGAAATTTGGGAAAAAACAAGGACGAGGAGGTCTCACCC---	436

Db	299	LysArgLysLysSerLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu	318
Qy	437	-----CAGGCACAGTACATTGACTCTTCCAGAACATCAAGTACAGC	478
Db	319	ArgAlaLysProProProAspGluPheValAspCysPheGlnLysPheLysHisGly	338
Qy	479	TTCACCTCTCGGAAGCTGGCCACCTGGCTGAAGGAGACAAAGTCCCTCAGCTCGTA	538
Db	339	PheAsnLeuLeuAlaLysLeuLysSerHisIleGlnAsnProSerAlaSerAspLeuVal	358
Qy	539	CACATCTCTTCAAGTCCCTGAACCTTCTCTGGCCAGCTGCCCTGAGGCTGGCCCTACA	598
Db	359	HisPheLeuPheThrProLeuAsnMetValGlnAlaThrGlyGlyProGluLeuAla	378
Qy	599	GCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTACAACCTGCCTACAGTCTGTCTA	658
Db	379	SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla	398
Qy	659	AGCCACCTGAGAGTAACTTTTGGATGGGTGGCCCGCAGCTGGACCACTAGCCGGGCC	718
Db	399	ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla	418
Qy	719	GACTGGACAGCGGATGAGCCCTG---CCCTACCAACCCACATCTCTCAGATGAGTGC	775
Db	419	GluTrpProLysGluGlnPheIleProTrpValProArgPheArgAsnGlyTrpGlu	438
Qy	776	CTTCCA-----GAGCCCTCCAGCCCAAGCACCTTAGGATACCAAGAC	817
Db	439	ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu	456
Qy	818	CCTGTTTCCCTTCGGCGGGGAAGTCAT---AGGTTAGGAGCAGCACCTCACATTTCTCTCAG	874
Db	457	AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr	476
Qy	875	GAGAAGACACACACCATGACCTCAGCCCTGGGAC-----	910
Db	477	GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerMet	496
Qy	911	-----CCC	913
Db	497	TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro	516
Qy	914	AAC-----TCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTCAAAATGCA	964
Db	517	AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys	536
Qy	965	GCTGTGTACAGTGTGAAGCTAGGAACCCACGGGAACCTGACTGTGTCCAGGAGAGAG	1024
Db	537	SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal	556
Qy	1025	CTGAGGTTCTGGACACAGCGGTGGTGGTGAAGATGAGGAGGAGGAGGAGGAGG	1084
Db	557	LeuGluIleLeuAspArgArgGlnTrpLysValArgAsnAlaSerGlyAspSer	576
Qy	1085	GGCTACATTCCAAGCAACATCTGGAG-----	1111
Db	577	GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg	596
Qy	1111	-----	1111
Db	597	AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer	616
Qy	1112	-----CCCCCTA	1117
Db	617	AlaAspThrProSerAlaProSerProProThrProAlaProValProValProLeu	636
Qy	1118	CAGCCGGGACCCCTGGG-----	1135
Db	637	ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln	656
Qy	1135	-----	1135
Db	657	AsnSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln	676

QY 1135 ----- 1135
Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
QY 1136 ----ACCCAGGGCAGTCACCCCTCTCGG-----GTTCCA 1165
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
QY 1166 ATGCTTCGACTTAGC-----TCGAGGCCTGAAGAGTTCACAGACTGGCTGCAGGCAGAG 1219
Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValIlyThrTrpLeuGlnSerLys 736
QY 1220 AACTTCTCCACTGCCACGCTGAGGACATTGGTCTCCTCAGCGGGACCGACTACTTTCGC 1279
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlnYalaGlnLeuPheSer 756
QY 1280 ATAAGACCTGGGAGCTACAGATGCTATGCTCCACAGGAGGCCCGCCAGTAATCTGTGCCCGG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValCysPro--GluGlyAlaArgValPheAsnGln 775
QY 1340 CTGGAGGCTGTCAGAGGATGCTG 1363
Db 776 IleThrValGlnLysAlaAlaLeu 783

RESULT 3

PCT-US93-07996-4

; Sequence 4. Application PC/TUS9307996

; GENERAL INFORMATION:

; APPLICANT: The Government of the United States, as represented by the
; APPLICANT: Secretary of Health and Human Services

; TITLE OF INVENTION: Substrate of the Epidermal Growth Factor Kinase

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe, Martens, Olson & Bear

; STREET: 620 Newport Center Drive, 16th Floor

; CITY: Newport Beach

; STATE: CA

; COUNTRY: USA

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/07996

; FILING DATE: 19930825

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 821 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

PCT-US93-07996-4

Alignment Scores:

Pred. No.:	4.16e-39	Length:	821
Score:	558.50	Matches:	158
Percent Similarity:	38.82%	Conservative:	78
Best Local Similarity:	25.98%	Mismatches:	171
Query Match:	17.86%	Indels:	201
DB:	5	Gaps:	17

US-09-762-021A-1 (1-1710) x PCT-US93-07996-4 (1-821)

QY 101 GGGCCTGCTATGAAGGCGGCTCCCTATGGAGCAGGCACGCTATCTGGAGCGCGGGATC 160

Db 190 GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209

QY 161 CCTCCAGAACAGCCCCACACAGAGACCCTTAGACACAGCCTCCACCATTCCTCCCAAGGCC 220
Db 210 ProProPro-----ProArgAlaProAlaPro 218
QY 221 CTGCCACGCCACACC-----AGTGCC 241
Db 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaAlaTrpSerAla 238
QY 242 CGAGAACAAGTGCCTTTACTCTGCTCCTCAAGCGCTCCTCTTCCCTCCCGAGGAC--- 298
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluGluThr 258
QY 299 CCAGAG-----AGGAGCAGAGAAAGTCTGAACCATGTCTTAAG 337
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
QY 338 GACATTGACCTTTTCATGGGAAGCTGGAGAAGGCCAG-----GCA 379
Db 279 AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
QY 380 AAGACCAGCAGGAAGAATAATTGGGAAAAAACAAGGACGAGGAGGTCTCACCC--- 436
Db 299 LysArgLysLysSerLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
QY 437 -----CAGGCACAGTACATTGACTCTCCAGAAGATCAAGTACAGC 478
Db 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
QY 479 TTCACCTCTCTGGGAAGCTGGCCACCTGGCTGAAGGACAGTCCCTCAGCTCGTA 538
Db 339 PheAsnLeuLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
QY 539 CATCTCTCTCAAGTCCCTGAACCTTCATCCTGGCCAGGTGCCCTGAGGCTGGCTAGCA 598
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyClyProGluLeuAla 378
QY 599 GCCCAAGTGATCTACCCCTCTCCTCACCCTTAAGCTATCAACCTGTACAGCTCTGTCTA 658
Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
QY 659 AGCCCACTGAGAGTAACCTTTGGATGGGTTGGCCCGCCAGCTGGACACTACCGCGGCC 718
Db 399 ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
QY 719 GACTGGACAGGCGATGAGCCCTG---CCCTACCAACCCACATCTCAGACTCTGCGCAA 775
Db 419 GluTrpProLysGluGlnPheIleProTyrValProArgPheArgAsnGlyTrpGlu 438
QY 776 CTCCA-----GAGCCCTCCAGCCCAAGCACCCCTAGATACACGAGAC 817
Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
QY 818 CTGTTTCCCTTCGGCGGGGAAGTCAT---AGTTAGGAGGACACCTCACACTTCTCTCAG 874
Db 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
QY 875 GAGAAGACACACAACCATCATCCTCAGCTGGGAC----- 910
Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerSerMet 496
QY 911 -----CCC 913
Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
QY 914 AAC-----TCAGGCGCCTCCAGCCCCAACCTGCCCGCCAGCCCTGAAATGCAA 964
Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
QY 965 GTCTGTACGAGTTGAAGCTAGGAACCCACGGAACCTGACTGTGTGTCCAGGAGAGAG 1024
Db 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal 556
QY 1025 CTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGAAGAATGAGCGGACCGAGC 1084

```
Db 557 LeuGluIleLeuAspArgArgGlnTrpTrpLysValArgAsnAlaSerGlyAspSer 576
Qy 1085 GGCTACATTCACAGCAACATCTGGAG-
Db 577 GlyPheValProAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
Qy 1111 -----
Db 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
Qy 1112 -----CCCTTA 1117
Db 617 AlaAspThrProSerAlaProSerProProProThrProAlaProValProValProLeu 636
Qy 1118 CAGCCGGGGACCCCTGGG-----
Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
Qy 1135 -----
Db 657 AsnSerSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
Qy 1135 -----
Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
Qy 1136 ---ACCAGGGCCAGTCACCTCTCGG-----GTTCCA 1165
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
Qy 1166 ATGTTCTCGACTTAGC-----TCGAGCCCTGAAGAGCTCACAGACTGGCTGCAGGCAGAG 1219
Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
Qy 1220 AACTTCTCACTGCCAGGTGAGACACTTGGGTCCCTGACGGGAGCCAGCTACTTCGC 1279
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
Qy 1280 ATAAGACCTGGGAGCTACAGACTATGTCCACAGGAGGCCACCACCACTCTCTCCCG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValLysPro---GluGlyAlaArgValPheAsnGln 775
Qy 1340 CTGGAGGCTGCAGAGGATGCTG 1363
Db 776 IleThrValGlnLysAlaAlaLeu 783
```

RESULT 4

```
US-08-931-820-4
; Sequence 4, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patencin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type III
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1055
; OTHER INFORMATION: /label= Modified
; OTHER INFORMATION: /note= "Ala may be Pro"
US-08-931-820-4

Alignment Scores:
Pred. No.: 9,01e-09 Length: 1057
Score: 200,50 Matches: 157
Percent Similarity: 32,01% Conservative: 37
Best Local Similarity: 25,91% Mismatches: 224
Query Match: 6,41% Indels: 188
DB: 3 Gaps: 34

US-09-762-021A-1 (1-1710) x US-08-931-820-4 (1-1057)
Qy 10 ACTGAACACAGCCTGCAGAA---GGCTCTGGAGGAGAGCTGGAGCAAGACCTCGACT 66
Db 498 ThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAla 517
Qy 67 TGGAGGCTTTCAGCCAGCCAGGACAGATGGAGGGGCTGCTATGGAAGGCCCTCCC 126
Db 518 GlyAlaProGlyAlaProGlyGlyLysAspAlaGlyAlaProGlyGluArgGlyPro 537
Qy 127 TATGGA-----GCAGGCACCTATCTGAGCCGGGATCCTCCAGAACAGGCC 174
Db 538 ProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProGly----- 554
Qy 175 CCACAGAGACCTAGA-----GCACAGCTCTCCACATCCCAAGGCCCTCCACG 228
Db 555 ---ProGluGlyGlyLysGlyAlaAlaGlyProProGlyProGlyAlaAlaGlyThr 573
Qy 229 CCACACAGTGCCTGAGAACCAAGTGCCTTACTCTCCTCCCAAGGCGTCTCTTC 288
Db 574 Pro-----GlyLeuGlnGly 578
Qy 289 CCCGAGGACCCAGAGAGGAGAGAGTGTGAACCATGTCTTAAGGACATTTGAGCT 348
Db 579 MetProGlyGluArgGlyGlyLeuGlySer-----ProGlyProLysGlyAsp----- 594
Qy 349 GTTCATGGGAAGCTGAGAGAGGCCAGGCAAGAACACAGCAGGAGAGAAATTTGGAA 408
Db 595 ---LysGlyGluProGly---GlyProGlyAlaAspGlyVal-----Pro 607
Qy 409 AAAAAACAAGGACCA---GGAGGTCTCACCCAGGCACAGTACATTGACTGCTTCCAGAA 465
Db 608 GlyLysAspGlyProArgGlyProThrGlyPro----- 618
Qy 466 GATCAAGTACAGCTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAGTGC 525
Db 619 -----IleGlyProProGlyProAlaGlyGln-----ProGlyAspLysGly 632
Qy 526 CCCTGAGCTCGTACACATCTCTTCAAGTCCCTGAATTCATCTCTGCGCAGGTG----CCC 582
Db 633 Gluglygly-----AlaProGlyLeu---ProGlyIleAlaGlyPro 645
Qy 583 TGAGGCTGGCTAGCAGCCCAAGTATCTACCCCTCTCACCCCTCAAGGATATCAACCT 642
Db 646 ArgGlySerProGlyGluArgGlyGluThrGlyProGlyProGlyProAlaGlyPheProGly 665
Qy 643 GCTACAGTCTCTTAAGCCACCTAGAGAGTAACTTTTGGATGGGTG----- 691
Db 666 AlaProGlyGlnAsn-GlyGluProGlyGlyLysGlyGluArgGlyAlaProGlyGluLys 685
Qy 692 -----GGCCACGCTGGACCACTAGCGGGCCGAGTGGACAGCGCATGAGCCCT 741
Db 685 sGlyGluGlyPro-ProGly---ValAlaGlyProProGlyGlySerGlyProAlaG 704
```

QY 742 GGCCTACCAACCCACATTTCTCAGATGACTGGCACTTCAGAGCCCT----- 788
Db 704 LyProProGlyProGlnGly---ValLysGlyGluArgGlySerProGlyGlyProGlyA 723
QY 789 -----CCAGCCAAAGCAGCCCTTAGGATACCA 813
Db 723 laAlaGlyPheProGlyAlaArgGlyLeuProGlyPro-ProGlySerAsnGlyAsnPro 742
QY 814 GGACCTGTTTCCCTTCGGCGGGGAAGTCATAGTTAGGGACCACTCACACTTTCCCTCA 873
Db 743 GlyProProGlyProSerGly-----SerPro 751
QY 874 GGAGAAGACACACAACTGACCTCA-----GCC 903
Db 752 GlyLysAspGlyProProGlyProAlaGlyAsnThrGlyAlaProGlySerProGlyVal 771
QY 904 TGGGAGACCCCACTCCAGGCCCTCCAGCCC-----CAACCC 939
Db 772 SerGlyProLysGlyAspAlaGlyGlnProGlyGlyLysGlySerProGlyAlaGlnGly 791
QY 940 TGCCAGCCAGCCCTCAAAATCCAACTCTTGTACGAGTTTGAAGCTTAGCTAGGAACCCAGGGA 999
Db 792 ProProGlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAla 811
QY 1000 ACTGACTGTGGTCCAGGGAGAGAGCTGGAGCTTCTGGACCAACAGCAAGCGGTGGTGGCT 1059
Db 812 GlyProProGlyMetProGlyProArgGlySerProGlyProGlnGly-----ValLys 829
QY 1060 GTTGAAGATGAGCC-----GGGACGGAGCGGCTACATTCCAAGCAACATCCTGGAGCC 1113
Db 830 GlyGluSer---GlyLysProGlyAlaAsnGlyLeuSerGlyGluArgGlyProProGly 848
QY 1114 CCT-----ACAGCGGGGAGCC-----TGGAGCCCA 1140
Db 849 ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGluProGlyArgAspGlyAsnPro 868
QY 1141 GGCCAGTCCACCTCTCGGCTTCCAACTGCTTCGACTTAGCTCGAGCGCTGAAGAGCTCAC 1200
Db 869 GlySerAspGlyLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGlu 888
QY 1201 AGACTGGCTGCAGGCGAGAGAACTTCTCCACTGCCAGTCCAGGCTGGAGACACTTGGCTCCCTGAC 1260
Db 889 -----Asn 889
QY 1261 GGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGAGGC 1320
Db 890 GlySerPro-----GlyAlaProGlyAlaProGlyHisProGly 902
QY 1321 CCACCAATCTGTCCGCTGGAGCTGTCCAGAGGATGCTGGGATAAG----- 1371
Db 903 ProProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluSerGlyProAla 922
QY 1372 ----CCCTTAGGACCCAGCTTACAGACCTCCAAAGACCGCCCGCTGTATGCAAGATGGC 1428
Db 923 GlyProAlaGlyAlaProGlyProAlaGlySerArgGlyAlaProGlyProGlnGlyPro 942
QY 1429 AGATCTGATACCCATAGAGCCCGGAGAAATTCCTCTTCTGGATCCAG----- 1476
Db 943 ArgGlyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGly 962
QY 1477 TTTGACGAAACCC-----ACACCCAGCTCACACAGCAAAACAAATGGACA 1524
Db 963 PheProGlyAsnProGlyAlaProGlySerPro----- 973
QY 1525 GGGCCAGAGGCTGAAGCAACAGTGT-----CCTTCTGCTGT 1563
Db 974 GlyProAlaGlyGlnGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyPro 993
QY 1564 GTTGGAGCCTCC 1575
Db 994 ValGlyProSer 997

RESULT 5
US-08-963-825-21
: Sequence 21, Application US/08963825
: Patent No. 6110689
: GENERAL INFORMATION:
: APPLICANT: Qvist, Per
: APPLICANT: Bonde, Martin
: TITLE OF INVENTION: A Method for Assaying Collagen Fragments
: TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
: TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
: TITLE OF INVENTION: Disorders Associated with the Metabolism of
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Darby & Darby PC
: STREET: 805 Third Avenue
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/963.825
: FILING DATE:
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/187.319
: FILING DATE: 21-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Gogoris, Adda C
: REGISTRATION NUMBER: 29,714
: REFERENCE/DOCKET NUMBER: 4305/08701
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-527-7700
: TELEFAX: 212-753-6237
: TELEX: 236687
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1078 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21
Alignment Scores:
Pred. No.: 5.78e-08 Length: 1078
Score: 191.00 Matches: 147
Percent Similarity: 29.11% Conservative: 37
Best Local Similarity: 23.26% Mismatches: 229
Query Match: 6.11% Indels: 219
DB: 3 Gaps: 26
US-09-762-021a-1 (1-1710) x US-08-963-825-21 (1-1078)
QY 7 GCGACTGAAGACCGCTCGAGAGGCTCTGGAGGAAGCTGGAGCAAGACCTCGACT 66
Db 334 AlaAlaGlyGluArgGlyAlaLeuGlySerArgGlyProAlaGly----- 348
QY 67 TGGAGCCCTTACGAGCCAGGACAGATGGAGGGGCGCTGCTATGGAAAGCCGCTCC 126
Db 349 -----ProAsnGlyIleProGlyGlu---LysGlyProAlaGlyGluArgGlyAlaPro 365
QY 127 TATGAGGAGGACCGCTATCTGGAGCCCGGGATCTCCAGAACAGCCCGCCAGGAGAC 186
Db 366 GlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGly 385


```

; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogotis, Adda C
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 4305/08701
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-527-7700
; TELEFAX: 212-753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1078 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: COLLAGEN ALPHA 1 (III)
US-09-500-811-21

Alignment Scores:
Pred. No.: 5,78e-08 Length: 1078
Score: 191.00 Matches: 147
Percent Similarity: 29.11% Conservative: 37
Best Local Similarity: 23.26% Mismatches: 229
Query Match: 6.11% Indels: 219
DB: 4 Gaps: 26

US-09-762-021a-1 (1-1710) x US-09-500-811-21 (1-1078)
QY 7 GCGACTGAAGACACGCTGCAAGAGCTGTGGAGGAGCTGGAGCAAGACCTCGACT 66
Db 334 AlaAlaGlyGluArgGlyAlaLeuGlySerArgGlyProAlaGly----- 348
QY 67 TGGAGCCTTCAGCAGGCGGACAGATGAGAGGGGCGCTGTATGGAAGCCGCTCC 126
Db 349 -----ProAsnGlyIleProGlyGlu-----LysGlyProAlaGlyGluArgGlyAlaPro 365
QY 127 TATGAGCAGGACGCTACTGTGAGCGGGGATCTCCAGAACACGCCACACGAGGAC 186
Db 366 GlyProAlaGlyProArgGlyAlaGlyGluProGlyArgAspGlyValProGlyGly 385
QY 187 CCTAGACACAGCCTCCACCATCCCAAGAGCCCTGCCACGCCACACAGTGCCTCCGAGA 246
Db 386 ProGlyMetArg-----GlyMetProGlySerProGlyGlyProGlySerAspGlyLys 403
QY 247 ACCAAGTGCCTTACTCTGCTCCCAAGGCGGCTCTTCCCGGAGGACCCAGAGAG 306
Db 404 ProGlyProProGlySerGlnGlyGluSerGlyArgProGlyProGlyProSerGly 423
QY 307 GGACGAGGAAGT-----GCTGAACCATGTCTTAAGGACATTGACCTGTCAT 354
Db 424 ProArgGlyGlnProGlyValMetGlyPheProGlyProGlyProGlyLysAspGlyAlaPro 443
QY 355 GGGAAAGCTGGAGAA-----GGCCCCAGGCAAGACAGCAGGAAAGAAATTTGGGAA 408
Db 444 GlyLysAsnGlyGluArgGlyGlyProGly----- 453
QY 409 AAAAAACAGGACGAGGAGGTCTCACCCAGGCACAGTACATGACTGCTTCCAGAAAGAT 468
Db 454 -----GlyProGly-----ProGlnGly 459
QY 469 CAAGTACAGCTTCAACCTCTCGGAAGGCTGCCACCTGGCTGAAGAGACAAAGTGCCCC 528
Db 460 -----ProProGlyLysAsnGlyGluTyrGly----- 468
QY 529 TGAGCTCGTACATCTCTTCAAGTCCCTGAACCTTCATCTGGCAGGTGCGCTGAGGC 588
Db 469 -----ProGlnGlyProProGly-----ProThrGly 477
QY 589 TGGCCTAGCAGCCCAAGTATCTACCCCTCTCACCCTTAAGATATCAACCTCGTACA 648
Db 589 -----ProGlnGlyProProGlyAlaGlnGlyProProGlyAlaProGly 798

```

```

Db 478 ProGlyGlyAspLysGlyAspThrGlyPro-ArgGlyProGlnGlyLeuGlnGlyLeuPr 497
QY 649 GTCCTGTCTAAGCCCAACCTGAGAGTAAC----- 676
Db 497 OGlyThrGlyGlyProProGlyGluAsnGlnLysProGlyGluProGlyProLysGlyCl 517
QY 676 ----- 676
Db 517 uAlaGlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGluArgG1 537
QY 677 -----CTTTGGATGGGTTGGCCGACGCTGACAC 708
Db 537 yProProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro-PropGlyPro- 556
QY 709 TAGCGGGCGGCGACTGACAGGCGATGAGCCCTGCCCTACCAACCAACATCTCAGATGA 768
Db 557 --GluGlyGlyLysGlyAlaAlaGlyProProGlyPro-----ProGlyAlaAlaGlyT 574
QY 769 CTGGCAACTTCCAGAGCCCTCCAGCCAAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCT 828
Db 574 hrProGlyLeuGlnGlyMetPro----- 581
QY 829 TCGCGCGGGAAGTCAATAGTTTAGGAGGACACCTCACACTTCTCAGGAGAGACACACNA 888
Db 582 -----GlyGluArgGlyGlyLeuGlySerProGlyPro----- 592
QY 889 CCATGACCTCAGCCTGGGACCCCA-----ACTCCAGGCCCTC 927
Db 593 -----LysGlyAspLysGlyGluProGlyGlyProGlyAlaAspGlyValProGlyLysA 611
QY 928 CAGCCCCAAACCTGCCAGCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCTAG 987
Db 611 spGlyProArgGlyProThrGlyPro-----IleG 621
QY 988 GAACCCAGGGAAGTACTGTGGTCCAGGAGAGAGAGCTGGAGGTTCTGGACCACAGCNA 1047
Db 621 ly-ProProGlyProAlaGlyGlnProGlyAspLysGlyGluGlyAlaProGlyLeu 640
QY 1048 GCGGTGGTGGCTGTGTAAGATGAGGCGGAGCGGCTACATTCCAAGCAACATCCT 1107
Db 641 ProGlyIleAlaGlyProArgGlySerProGlyGluArg----- 653
QY 1108 GGAGCCCTACAGCGGGAGCCCTGGGACCCAGGCGGACGTCACCTCTCGGTTTCCAAT 1167
Db 654 -----GlyGluThrGlyProGlyProAlaGlyPheProGlyAlaPro 668
QY 1168 GCTTCGACTTAGCTCGAGCCCTGAAGAGTCAACACTGGCTGCGAGGAGAGAACTTCTC 1227
Db 669 GlyGlnAsnGlyGluProGlyGlyGlyGluArgGlyAlaProGlyGlyLys----- 686
QY 1228 CACTGCCACGGTGAAGACACTTGGGTCCCTGACGGGAGCCAGCTACTTCGCATAAGACC 1287
Db 687 -----GlyGlu-----GlyGlyProProGlyValAlaValProPro 698
QY 1288 TGGGAGCTACAGATGCT-----ATGTC 1311
Db 699 GlyGlySerGlyProAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySer 718
QY 1312 ACAGAGGAGCC----- 1323
Db 719 ProGlyGlyProGlyAlaAlaGlyPheProGlyAlaArgLysLeuProGlyProGly 738
QY 1324 -----ACGAATCTTCTCCCGCTGAGGCTGTCAGAAGGATGCTGGGATAGCCCTTA 1377
Db 739 SerAsnGlyAsnProGlyProProGlyProSerGlySerProGlyLysAspGlyProPro 758
QY 1378 GGCACCACTTAGACACCTCCAAAGAACCCAGGCCGCTGATGCAAGATGGCAGATCTGAT 1437
Db 759 GlyProAlaGlyAsnThrGlyAlaProGlySerProGlyValSerGlyProLysGlyAsp 778
QY 1438 ACCCATAGACCCCGAGAAATTCCTTCTGATCCAGTTTGCACCAAC-----CCCACA 1494
Db 779 AlaGlyGlnProGlyGlyLysGlySerProGlyAlaGlnGlyProProGlyAlaProGly 798

```



```
Db 454 -----GlyProGly-----ProGlnGly 459
Qy 469 CAAGTACAGCTTCAACCTCTCGGAAGCGTGGCCACCTGGCTGAAGGAGACAAAGTGGCCCC 528
Db 460 -----ProProGlyLysAsnGlyGluThrGly----- 468
Qy 529 TGAGCTCGTACACATCTCTTCAAGTCCCTGAACCTTCATCTGCGCAGGTGCCGTAGGC 588
Db 469 -----ProGlnGlyProProGly-----ProThrGly 477
Qy 589 TGGCTAGCAGCCCAAGTATCATCCCTCCCTCCTCACCCTTAACCTATCAACCTGCTACA 648
Db 478 ProGlyGlyAspLysGlyAspThrGlyPro-ArgGlyProGlnGlyLeuGlnGlyLeuPr 497
Qy 649 GTCCTGCTTAAGCCCACTCGAGATAAC----- 676
Db 497 oGlyThrGlyGlyProProGlyGluAsnGluLysProGlyGluProGlyProLysGlyG 517
Qy 676 ----- 676
Db 517 uAlaGlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGluArgG 537
Qy 677 -----CTTTGGATGGGTGGTGGCCAGCCGCTGAGCCAC 708
Db 537 yProProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro-ProGlyPro- 556
Qy 709 TAGCCGGCCGACTGGACAGCGATGAGCCCTCCCTACCACCCACATCTCAGATGA 768
Db 557 --GluGlyGlyLysGlyAlaGlyProProGlyPro-----ProGlyAlaGlyT 574
Qy 769 CTGGCAACTCCAGAGCCCTCCAGCCCAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCT 828
Db 574 hrProGlyLeuGlnGlyMetPro----- 581
Qy 829 TCGCGGGGAGTCAATAGTTAGGAGCACTCACATCTTCTCAGGAGAGACACAA 888
Db 582 -----GlyGluArgGlyGlyLeuGlySerProGlyPro----- 592
Qy 889 CCATGACCCCTCAGCTGGGACCCCA-----ACTCCAGGCCCTC 927
Db 593 -----LysGlyAspLysGlyGluProGlyProGlyAlaAspGlyAlaProGlyLys 611
Qy 928 CAGCCCAAACTCCAGCAGCGAGCCCTGAAATGCAAGTCTTGACGAGTTGAAGCTAG 987
Db 611 spGlyProArgGlyProThrGlyPro-----1166 621
Qy 988 GAACCCACGGGAAGTACTGTGTCCAGGAGAGAGCTGGAGTTCTGGACCAACAGCAA 1047
Db 621 ly-ProProGlyProAlaGlyGlnProGlyAspLysGlyGlyGlyAlaProGlyLeu 640
Qy 1048 GCGGTGTGTGTGAAGATAGGCGGAGCGGCTACATTCACAGCAACATCCT 1107
Db 641 ProGlyLeuAlaGlyProArgGlySerProGlyGluArg----- 653
Qy 1108 GGAGCCCTACAGCCGGGACCCCTGGGACCCAGGCGGAGTCCCTCGGTTCCAAT 1167
Db 654 -----GlyGluThrGlyProProGlyProAlaGlyPheProGlyAlaPro 668
Qy 1168 GCTTTCGACTTAGCTCGAGGCTTGAAGAGGTACAGACTGGCTGCAGGAGCACTTCTC 1227
Db 669 GlyGlnAsnGlyGluProGlyGlyLysGlyGlyArgGlyAlaProGlyGlyLys----- 686
Qy 1228 CACTGCCAGCGTGAAGACACTGGTCCCTGACGGGAGCCAGTACTTCGCATAAGACC 1287
Db 687 -----GlyGlu-----GlyGlyProGlyValAlaValProPro 698
Qy 1288 TGGGAGCTACAGATGCT-----ATGTC 1311
Db 699 GlyGlySerGlyProAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySer 718
Qy 1312 ACAGGAGGCC----- 1323
```

```
Db 719 ProGlyGlyProGlyAlaAlaGlyPheProGlyAlaArgGlyLeuProGlyProProGly 738
Qy 1324 -----ACGAATCTCTCCCGGCTGGAGGCTGTGTCAGAGGATGCTGGGATAGCCCTTA 1377
Db 739 SerAsnGlyAsnProGlyProProGlyProSerGlySerProGlyLysAspGlyProPro 758
Qy 1378 GGCACACAGCTTAGACACTCCAAGAACACAGCCCGCTGATGCAAGATGCGAGATCTGAT 1437
Db 759 GlyProAlaGlyAsnThrGlyAlaProGlySerProGlyValSerGlyProLysGlyAsp 778
Qy 1438 ACCATTAGAGCCCGAGAATTCCTCTTCTGGATCCAGTTTCAGCAAAAC---CCACA 1494
Db 779 AlaGlyGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGlyAlaProGly 798
Qy 1495 CCC-----CAGCTCACACAGCAAAACAATGGACAGCCCGAGGCT----- 1536
Db 799 ProLeuGlyLeuAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetPro 818
Qy 1537 -----GAAGCAACAGTGTCCCTTCTGGTGTGTTGGAGCTCCCGAGTAAACACCTAT 1590
Db 819 GlyProArgGlySerProGlyProGlnGlyValLysGlyGluSer----- 833
Qy 1591 TTATTTTACCTCTTCCCAACCTGGAGCA 1620
Db 834 -----GlyLysProGlyAla 838

RESULT 9
US-08-642-255-52
; Sequence 52, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BTR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-52

Alignment Scores: 2.85e-07 Length: 561
Pred. No.: 181.50 Matches: 131
Score: 31.03% Conservative: 13
Percent Similarity:
```

Best Local Similarity: 28.23% Mismatches: 204
Query Match: 5.80% Indels: 116
DB: 1 Gaps: 27

US-09-762-021a-1 (1-1710) x US-08-642-255-52 (1-561)

QY 25 GCAGAGGCTCTGGAGAGAGCTGGAGCAAGACCTCGACTTGGAGCGCTTCAGCC--- 81
DB 188 AlaProGlyProAlaGlyProGlyAlaProGlyProAlaGlyProGlyAlaPro 207
QY 82 AGCCAGGACAGATGAGGGGCGCTGCTATGGAAGGCGCTCCATATGGAGCA---GGC 138
DB 208 GlyProAlaGlyProProGlyAla---ProGlyProAlaGlyProProGlyAlaProGly 226
QY 139 ACGTATCTGGAGCGGGGATCCCTCCAGAACAGCCACAGCCAGGAGCCCT---AGAGCA 195
DB 227 ProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla 246
QY 196 CAGCCTCCACCATCCCCAAGGCCCTGCCAGCCACACCAAGTGGCCGAGAACCAAGTGC 255
DB 247 GlyProProGlyAlaPro---GlyProAlaGlyProProGlyAlaProGly----- 262
QY 256 CTTTACTCTGCTCTCCCAAGCGGTCTCTTCCCGGAGGACCCAGAGAGGAGGAGCA 315
DB 263 -----ProAlaGlyProProGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly 280
QY 316 AGTGCTGAACCA-----TGCTTAAGGACATTGAGCTGTTTCATGGGAAAGCT 363
DB 281 ProAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAla 300
QY 364 GGAGAGGCCCGGCAAGACAGCAGGAGGAAGAAGAAATTTGGGAAAAAACAAGGACCA 423
DB 301 Gly-----ProLysGlyAlaPro 306
QY 424 GGGAGTCTACCCAGGCACAGTACATTGACTGCTTCAGAGAGTCAAGTACAGCTTCAA 483
DB 307 GlyProAlaGlyPro-----ProGlyAlaProGlyProAlaGly 319
QY 484 CTCTCTGGGAAGCTGGCCACCTG-----GCTGAAGGAGACAAAGTGGCCCTGA 531
DB 320 ProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProPro 339
QY 532 GCTCGTACACATCTCTTCAAGTCCCTGAACCTCATCTGCGCCAGGTGCCCTGAGGCTGG 591
DB 340 GlyAlaProGlyProAlaGlyProProGlyAla---ProGly-----ProAlaGlyPro 356
QY 592 CCTAGCAGCCCAAGTATCTCACCCCTCTCACCCCTTAAGCTATCAACCTCTCTACAGTC 651
DB 357 ProGlyAlaProGlyProAlaGlyProProGly-----AlaProGlyProAlaGlyPro 374
QY 652 CTGCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGCCCGCAGCTGGACCA--- 707
DB 375 ProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla---GlyProProG 394
QY 708 --CTAGCCGGCGAGTGGACAGGCGATGAGCCCTGCCCTACCAACCCACATCTTCAGA 765
DB 394 lYAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGly---- 412
QY 766 TGACTGGCAACTCCAGAGCCCT---CCAGCCCAAGCACCCCTTAGGATACCCAGGACCCCTGT 822
DB 413 -----AlaProGlyProAlaGlyProPro---GlyAlaProGlyProAla 426
QY 823 TTCCCTTCGGCGGGGAGTCATAGGTTAGGGAGCACCTCACACTTTCCTCAGGAGAAC 882
DB 427 GlyProProGlyAlaProGly-----ProAlaGlyProPro 438
QY 883 ACACAACTATGACCTCAGCCCTGGGAGCCCACTCCAGGCGCTCCAGCCCAACACCTGC 942
DB 439 GlyAlaProGlyProAla-----GlyProPro---GlyAlaProGlyProAlaGlyPro 455
QY 943 CCAGCCAGCCCTGAANATCAAGTCTTGTACGAGTTTGAAGCTAGAACCCACGGGAAGCT 1002
DB 456 ProGlyAlaProGlyProAlaGly-----ProProGlyAla 467

QY 1003 GACTGTGCTCAGGAGAGAAAGCTGGAGGTTCTTGACACACAGCAAGCGGTGCTGCTGCT 1062
DB 468 Pro---GlyProAlaGlyProProGlyAlaProGlyPro-----AlaGly 481
QY 1063 GAAGAATGAGCGGGGAGCGGAGGCTACATCCCAAGCAACATCTCTGAGGAGCCCT---ACA 1119
DB 482 Pro-----ProGlyAlaProGlyPro 488
QY 1120 GCCGGGACCCCTGGGACCCAGGCGCCAGTACCCTCTCGGGTTCCAATGCTTCGACTTAG 1179
DB 489 AlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla--- 507
QY 1180 CTCGAGGCTGAAGAGTGCACAGAGTGGCTGCAGGAGCAAGAACTTCTCCACTGCCACGCT 1239
DB 508 -----GlyProGlyAlaProGly----- 514
QY 1240 GAGGACACTTGGTCCCTGACGGGAGGAGCCAGTACTTCGCATAAGACCTGGGAGCTACA 1299
DB 515 -----ProAlaGlyProProGly---AlaHisGlyProAlaGlyProLys 528
QY 1300 GATGCTATGTCACAGAGGAGGCGCCACGAATCTGTCCCGGCTGGAGGCTGTCAAGAGGAT 1359
DB 529 GlyAlaHisGlyProAlaGlyProLysGlyAlaMetAspProGlyArgTyroGlnLeuSer 548
QY 1360 GCTGGG 1365
DB 549 AlaGly 550
RESULT 10
US-09-219-849-3
; Sequence 3, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOUWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WESTEN, MARC W.T.
; APPLICANT: WIND, RICHELIE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE OF INVENTION: PREPARATION THEREOF
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-3

Alignment Scores:
Pred. No.: 4,04e-07 Length: 504
Score: 179.50 Matches: 130
Percent Similarity: 29.64% Conservative: 9
Best Local Similarity: 27.72% Mismatches: 184
Query Match: 5.74% Indels: 146
DB: 4 Gaps: 27

US-09-762-021a-1 (1-1710) x US-09-219-849-3 (1-504)

QY 25 GCAGAGGCTCTGGAGGAGAGAGCTGGAGCAAGACCTCGACTTGGAGCGCTTCAGCC--- 81
DB 155 AlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProGlyAlaPro 174

```
Qy 82 AGGCCAGACAGATGGAGGGCCCTGCTATGTGAAGGCCGCTCCTATGTGAGCA---GGC 138
Db 175 GlyProAlaGlyProProGlyAla---ProGlyProAlaGlyProProGlyAlaProGly 193
Qy 139 ACGTATCTGGAGCGGGGATCCCTCCAGACACAGCCACCAGAGGACCT---AGAGCA 195
Db 194 ProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla 213
Qy 196 CAGCCTCCACCATCCCAAGGCCCTCCACGACACACACAGTCCCGGAGAACCAAGTGC 255
Db 214 GlyProProGlyAlaPro---GlyProAlaGlyProProGlyAlaProGly----- 229
Qy 256 CTTTACTCTGCTCCTCCAAAGCGGTCTCTTCCCGAGGACCCAGAGGGAGGAGGA 315
Db 230 -----ProAlaGlyProProGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly 247
Qy 316 AGTGCTGACCA-----TGTCCTAAGGGAGCATTCAGCTGTTTCATGGGAAGCT 363
Db 248 ProAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAla 267
Qy 364 GGAGAAGGCCAGGCAAGACACGACGAGGAAGAAATTTGGGAAAAAACAAGGACCA 423
Db 268 Gly-----ProLysGlyAlaPro 273
Qy 424 GGGAGGTCTCACCAGGCACAGTACATTGCTGCTTCCAGAGATCAAGTACAGCTTCAA 483
Db 274 GlyProAlaGlyPro-----ProGlyAlaProGlyProAlaGly 286
Qy 484 CTTCTGGGAGGCTGGCCACCTG-----GCTGAAGGAGACAAAGTGCCCTCGA 531
Db 287 ProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProPro 306
Qy 532 GCTCGTACACATCTCTTCAAGTCCCTCAAGTTCATCTGCTGCCAGGTCCTGAGGCTGG 591
Db 307 GlyAlaProGlyProAlaGlyProProGlyAla---ProGly-----ProAlaGlyPro 323
Qy 592 CTTAGCAGCCCAAGTGTCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGTC 651
Db 324 ProGlyAlaProGlyProAlaGlyProProGly-----AlaProGlyProAlaGlyPro 341
Qy 652 CTGTCTAGCCCACTGAGAGTACCTTTGGATGGGTGGGCGCCAGCCGTCAGCA----- 707
Db 342 ProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla---GlyProProG 361
Qy 708 --CTAGCGCGCCGACTGGACGCGATGAGCCCTGCTTACCAACCCACATTCCTCAGA 765
Db 361 lyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGly---- 379
Qy 766 TGACTGGCACTTCCAGAGCCCT---CCAGCCAAAGCACCCTTAGGATACCAAGGACCTGT 822
Db 380 -----AlaProGlyProAlaGlyProPro-GlyAlaProGlyProAla 393
Qy 823 TTCCTTTGGCGGGAAGTCATAGTTAGGAGGACCTCACACTTCTCAGGAGAGAC 882
Db 394 GlyPro----- 395
Qy 883 ACACAACCATGACCTCAGCCTGGGAGCCCAACTCCAGGCGCTCCAGCCCAACCTGC 942
Db 396 -----ProGlyAlaProGlyProAlaGlyProProGlyAlaProGly----- 409
Qy 943 CCAGCCAGCCCTGAAATGCAAGTCTTTGTACAGTTTGAAGCTAGGAACCCACGCGGA 1002
Db 410 ProAlaGlyProProGlyAlaPro----- 417
Qy 1003 GACTGTGTCCAGGAGAGAGCTGGAGTTCTGGACCACAGCAAGCGGTGGTGGTGT 1062
Db 418 -----GlyProAlaGlyProProGlyAlaProGlyPro-----AlaGly 430
Qy 1063 GAAGAACTGAGCGGAGCGGCTACATTCCAAGCAACATCTCTGGAGCCCT---ACA 1119
Db 431 Pro-----ProGlyAlaProGlyPro 437
Qy 1120 GCCGGGACCCCTGGGACCCAGGCGCAGTCACCTCTCGGTTCCAAATGCTTCGACTTAG 1179
```

```
Db 438 AlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGly----- 454
Qy 1180 CTCGAGCGCTGAAGAGGTACACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCACGGT 1239
Db 455 -----ProAlaGly----- 457
Qy 1240 GAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACA 1299
Db 458 -----ProGlyAlaProGlyProAla-----GlyProPro 468
Qy 1300 GATGCTATGTCCACAGGAGGCCCCACGAAT---CCTGTCCCGGCTGGA-----GGCTGT 1350
Db 469 GlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAla 488
Qy 1351 CAGAAGGATGCTGGGATAG 1371
Db 489 HisGlyProAlaGlyProLys 495
RESULT 11
US-09-219-849-7
; Sequence 7, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 7
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-7
Alignment Scores:
Pred. No.: 5,61e-07 Length: 552
Score: 178.00 Matches: 153
Percent Similarity: 29.93% Conservative: 26
Best Local Similarity: 25.59% Mismatches: 264
Query Match: 5.69% Indels: 155
DB: 4 Gaps: 33
US-09-762-021a-1 (1-1710) x US-09-219-849-7 (1-552)
Qy 22 CTGCAAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCAGCC 81
Db 57 ProAlaGlyProGlyGlySerArgAspProGlyProGlyAlaGlnGlyProAlaGly 76
Qy 82 AGGCCAGACAGATGGAGGGCCCTGCTATGGAAGGCCGCTCCTATGGAGGAGGACG 141
Db 77 ProGlyGlySerArgAspProGlyProGlyAlaGlnGlyProAlaGlyProGlyGly 96
Qy 142 CTATCTGGAGCGGGGATCCCTCCAGAACAGCCCA----- 177
Db 97 SerArgAspProGlyProProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgAsp 116
Qy 178 CCAGGAGCCCTTAGAGCAGACGCTCCCAACCATCCCAAG-----GCCCTGCCAGC 228
```

```

Alignment Scores:
Pred. No.:          9.03e-07          Length:      1461
Score:             177.50           Matches:      163
Percent Similarity: 27.59%          Conservative: 29
Best Local Similarity: 23.42%       Mismatches:   222
Query Match:       5.68%            Indels:       282
DB:                4                Gaps:         33

US-09-762-021A-1 (1-1710) x US-09-585-887-9 (1-1461)

Qy      1  GGCAGAGCGACTGAAGACCAACGCGCTGCA-----GAAGGCTCTGGAGGAGAGAGCTGGAGCA 54
      |||  |||  ||| ||||| ||||| ||||| ||||| |||  |||||

```


Db 497 GlyValAlaGlyProLysGlyProAlaGlyGluArgGlySerProGlyProAlaGly--- 515
QY 55 AAGACCTCGACTTGGAGCCCTTCAGCCAGCCAGCAGATGGAGGGGCTGCTATGGA 114
Db 516 -----ProLysGlySerProGly-----GluAlaGlyArgProGly 527
QY 115 AAGCCCGCTCCCTATGGAGCGGCACGCTATCTCTGAGCC-----GGGATCCC 162
Db 528 GluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAsp 547
QY 163 TCCAGAACGCCCCACCA-----GAGACCCCTAGAGCACAGCCTCCACCATC 210
Db 548 GlyThrGlyProGlyProAlaGlyProAlaGlyAspGlyArgProGlyProPro----- 565
QY 211 CCCAAGGCCCTGCGCACGCCACACCTAGTCCCGAGACCAAGTGCCTTTACTCTGCTCC 270
Db 565 ----- 565
QY 271 TCCAAGCGGTCTCTTCCCGGAGGACCCAGAGAGGAGGAGGAGTGGTGA----- 324
Db 566 -----GlyProGlyAlaArgGlyGlnAlaGlyValMet 577
QY 325 -----CCATGCTCTAAGGACATGAGCTGTTCATCGGAAGCTGGAGAA----- 369
Db 578 GlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyGluArgGlyVal 597
QY 370 -----GGCCCA-----GGCAAGACCCAGCAGGAGGAAGAAATTTGG 405
Db 598 ProGlyProGlyAlaValAlaGlyProAlaGlyLysAspGlyGlu----- 612
QY 406 GAAAAAACAAGGACGAGGAGGTCTCACCAGCCAGCAGTACATTGACTGCTTCCAGAA 465
Db 613 AlaGlyAlaGlnGlyPro-----ProGly----- 620
QY 466 GATCAAGTACAGCTTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGAGACAAAGTGC 525
Db 621 -----ProAlaGlyProAlaGlyGluArgGlyGlnGlnGlyProAla 634
QY 526 CCGTGAGCTGTACACATCTCTTCAAGTCCCTGAAGTTCATCTGCGCCAGGTCCCTGA 585
Db 635 GlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProProGlyGluAla----- 652
QY 586 GGCTGGCTAGCAGC-----CCAAGTGTATCTCACCCCTCTCACCCCTAAAGCTAT 636
Db 653 GlyLysProGlyGlnGlnGlyValProGlyAspLeuGlyAlaProGlyPro----- 669
QY 637 CAACCTGCTACAGTCTGTCTTAACCCACCTGAGAGTAACCTTTGGATGG----- 687
Db 670 -----SerGlyAlaArgGlyGluArgGlyPheProGlyGluArgGly 683
QY 688 ---GTTGGGCCAGCTGGAGCCACTAGCGGCGGCGGAGTGGACAGCGCATGAGCC----- 740
Db 684 ValGlnGlyPro--ProGlyPro--AlaGlyProArgGlyAlaAsnGlyAlaProGlyA 702
QY 741 -----TGCCCTTACCACCCACAPCTCAG----- 764
Db 702 snAspGlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAlaProG 722
QY 764 ----- 764
Db 722 lyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyProLysGlyAspA 742
QY 765 -----A 765
Db 742 rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyL 762
QY 766 TGACTGGCAACTTCCAGAGCCCT---CCAGCCAAAGCACCCTTAGGATACCAGGACCCCTGT 822
Db 762 euThrGlyProIleGlyProProGlyProAlaGlyAlaProGlyAspLysGlyGluSerG 782
QY 823 TTCCCTTCGGC-----GGGAAGTCAATAGGTTAGGAGGACCTTCACACTTCTCAGGA 876
Db 782 lyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlaProGly----- 797

QY 877 GAAGACACACACCATGACCCCTCAGCCTGGGACCCCACTCCA----- 920
Db 798 -----AspArgGlyGluProGlyProGlyProAlaGlyP 810
QY 920 ----- 920
Db 810 heAlaGlyProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspAlaG 830
QY 921 -----GGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATGCA 963
Db 830 lyAlaLysGlyAspAlaGlyProProGlyPro--AlaGlyProAlaGlyProProGly--- 848
QY 964 AGTCTTCTACGAGTTTGAAGCTAGGAACCCAGCGGAAGTACTGTGCTCCAGGAGAGAA 1023
Db 849 -----ProIleGlyAsnValGlyAlaProGlyAlaLys 859
QY 1024 GCT-----GGAGTTCCTGGACACACAAAGCGGTG----- 1053
Db 860 GlyAlaArgGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArg 879
QY 1054 -----GTGCTGGT 1062
Db 880 ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGlyProAlaGly 899
QY 1063 GAAGAATGAGCGGGACCGAGCGG----- 1086
Db 900 LysGluGlyGlyLysGlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluVal 919
QY 1087 -----CTACATTCCAAGCAACATCCTGAGGCCCT---ACA 1119
Db 920 GlyProProGlyProProGlyProAlaGlyGluLysGlySerProGlyAlaAspGlyPro 939
QY 1120 GCGGGGACCCCTGGGACCCAGGCCAGTCACCTCTCGGGTTCATGCTTCCACTTAG 1179
Db 940 AlaGlyAlaProGlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGly 959
QY 1180 CTCGAGCCCTGAAGAGTTCACAGCTGGCTGCGAGCAGAGAACTTCTCCACTGCCACGGT 1239
Db 960 LeuProGlyGlnArgGlyGluArgGlyPheProGlyLeuPro----- 973
QY 1240 GAGGACACTTGGTCTCCTGACGGGAGCCAGCTACTTCGCATAAGACCTGGGAGCTACA 1299
Db 974 -----GlyProSerGlyGluProGlyLysGlnGlyProSer---GlyAlaSer 988
QY 1300 GATGCTATGTCCACAGGAGGCCCCACCAATCTCTCCGGCTGGAGCTGTCAAGAGAT 1359
Db 989 GlyGluArgGlyProProGlyProMetGlyProProGlyLeuAlaGlyProProGlyGlu 1008
QY 1360 GCTGGG---GATAGCCCTTAGGCACCCAGCTTAGACACCTCCAAGAACAGGCCCGCTG 1416
Db 1009 SerGlyArgGluGlyAlaProGlyAlaGlySerProGlyArgAspGlySerProGly 1028
QY 1417 ATCAAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAATTCCTTCTTGATCCAG 1476
Db 1029 AlaLysGlyAspArgGlyGluThrGlyProAlaGlyProProGlyAlaProGlyAlaPro 1048
QY 1477 TTGCAACAACCCACACCCCGCTC-----ACACGAGAAAACAACATGACAGGCCCA 1530
Db 1049 ValAlaProGlyProValGlyProAlaGlySerGlyAspArgGlyGluThrGlyPro 1068
QY 1531 GAGGCTCAAGCAACACAGTGTCTTCTGGCTGTGTGGAGCC 1572
Db 1069 AlaGlyProAlaGly-----ProValGlyProValGlyAla 1080

RESULT 13

US-09-289-578-9
: Sequence 9, Application US/09289578
: Patent No. 6428978
: GENERAL INFORMATION:
: APPLICANT: Olsen, David R
: APPLICANT: Chang, Robert
: APPLICANT: McMullin, Hugh

```
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/289,578
; CURRENT FILING DATE: 1999-04-10
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patencin ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-289-578-9

Alignment Scores:
Pred. No.:          9,03e-07      Length:      1461
Score:             177.50        Matches:    163
Percent Similarity: 27.59%       Conservative: 29
Best Local Similarity: 23.42%    Mismatches: 222
Query Match:       5.68%        Indels:     282
DB:                4           Gaps:       33

US-09-762-021a-1 (1-1710) x US-09-289-578-9 (1-1461)
QY 1  GCGAGCGGCTGAGACGAGCTGCA-----GAAGCTCTGGAGGAGAGCTGGAGCA 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 497 GlyValAlaGlyProLysGlyProAlaGlyGluArgGlySerProGlyProAlaGly--- 515
QY 55 AAGACTCGACTGGAGGCTTCAGCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 516 -----ProLysGlySerProGly-----GluAlaGlyArgProGly 527
QY 115 AAGGCGGCTCCCTATGGAGCGGACGCTATCTGGAGCC-----GGGATCCC 162
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 GluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAsp 547
QY 163 TCCAGNAGACCCCA-----GAGGACCTAGAGCAGACAGCCTCCACCATC 210
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 548 GlyLysThrGlyProGlyProAlaGlyGluAspGlyArgProGlyProPro----- 565
QY 211 CCCAAGGCGCTGCCAGCGCACAGCTGCCGAGAACCAAGTGCCTTTACTCTGCCTCC 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 ----- 565
QY 271 TCCAAGGCGGCTCTCTTCCCCGAGGAGCCAGAGAGGAGGAGGAGGAGGAGGAGGAG 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 -----GlyProGlyAlaArgGlyGluAlaGlyValMet 577
QY 325 -----CCATGTCCTAAGGACATTGACGTGTTTCATGGGAAAGCTGGAGAA----- 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 GlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyGluArgGlyVal 597
QY 370 -----GGCCCA-----GGCAAGACACAGGAGGAGGAGGAGGAGGAGGAGGAG 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 ProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGlu----- 612
QY 406 GAAAAAACAAGGACGAGGAGGAGTCTCACCAGGAGCAGTACATTGACTGCTTCCAGAA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 613 AlaGlyAlaGlnGlyPro-----ProGly----- 620
QY 466 GATCAAGTACAGCTCAACCTCTGGGAGGCTGGCCACCTGGCTGAGGAGACAGTGC 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 621 -----ProAlaGlyProAlaGlyGluArgGlyGluGlnGlyProAla 634
QY 526 CCTGTAGCTCGTACATCCTCTCAAGTCCCTGAATCTCACTCCCTGGCCAGGTGCCCTGA 585
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 635 GlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProProGlyGluAla----- 652
QY 586 GCGTGGCTTACGAGC-----CCAAGTGATCTACACCCCTCTCACCCCTCAAAGCTAT 636
```

```
Db 653 GlyLysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyPro----- 669
QY 637 CAACCTGCTACAGTCTCTTAAGCCACCTGAGAGTAACCTTTGGATGGG----- 687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 670 -----SerGlyAlaArgGlyGluArgGlyPheProGlyGluArgGly 683
QY 688 ---GTTGGGCCAGCTGGACACTAGCCGGCCGAGCTGGACAGCGATGAGCCCC--- 740
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 684 ValGlnGlyPro--ProGlyPro---AlaGlyProArgGlyAlaAsnGlyAlaProGlyA 702
QY 741 -----TCCCTACCACCCACATCTCTCAG----- 764
Db 702 snAspGlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAlaProG 722
QY 764 ----- 764
Db 722 lyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyProLysGlyAspA 742
QY 765 -----A 765
Db 742 rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyL 762
QY 766 TGACTGGCAACTTCCAGAGCCCT---CCAGCCACACCCCTTAGATACAGGACCCCTGT 822
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 762 euThrGlyProIleGlyProGlyProAlaGlyAlaProGlyAspLysGlyGluSerG 782
QY 823 TTCCCTTCGGC-----GGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGA 876
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 782 lyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlaProGly----- 797
QY 877 GAAGACACACAACCATGATGACCTCGGAGGAGCCCACTCCA----- 920
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 798 -----AspArgGlyGluProGlyProProGlyProAlaGlyP 810
QY 920 ----- 920
Db 810 heAlaGlyProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspAlaG 830
QY 921 -----GSCCTCCAGCCCAACCTCCCGAGCCAGCCCTGAAATGCA 963
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 lyAlaGlyAspAlaGlyProGlyPro--AlaGlyProAlaGlyProProGly--- 848
QY 964 AGTCTTGACGAGTTGAAGCTAGGAACCCAGCGGAGCTGCTGTGTCAGGGAGAGAA 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 -----ProIleGlyAsnValGlyAlaProGlyAlaLys 859
QY 1024 GCT-----GGAGGTTCTGGACCAAGCGGTG----- 1053
    ||| :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 860 GlyAlaArgGlySerAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArg 879
QY 1054 -----GTGGCTGT 1062
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 880 ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGlyProAlaGly 899
QY 1063 GAAAGTATGAGCGGAGCGGAGCGG----- 1086
    :||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 900 LysGluGlyLysGlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluVal 919
QY 1087 -----CTACATCCCAAGCAACATCTCGGAGCCCT---ACA 1119
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 920 GlyProProGlyProProGlyProAlaGlyGluLysGlySerProGlyAlaAspGlyPro 939
QY 1120 GCCGGGAGCCCTGGAGCCAGCGCTGCTGGGTTCAGTTCAGTCTCGACTTAG 1179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 940 AlaGlyAlaProGlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGly 959
QY 1180 CTCGAGGCTGAAGAGTGCACAGCTGGCTGCAGGAGAGAACTTCTCCACTGCCACGGT 1239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 960 LeuProGlyGlnArgGlyGluArgGlyPheProGlyLeuPro----- 973
QY 1240 GAGGACACTTGGTCCCTCGGAGGAGCAGCTACTTCCATTAAGACCTGGGAGCTACA 1299
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 974 -----GlyProSerGlyGluProGlyLysGlnGlyProSer----GlyAlaSer 988
Qy 1300 GATGCTATGTCACAGGAGGCCACGAATCCTGTCCCGGTGGAGCTGTCTCAGAAGGAT 1359
Db 989 GlyGluArgGlyProProGlyProMetGlyProGlyLeuAlaGlyProGlyGlu 1008
Qy 1360 GCTGGG---GATNAGCCCTTAGGACACAGCTTAGACACCTCCAGAACAGGCCCGCTG 1416
Db 1009 SerGlyArgGluGlyAlaProGlyAlaGlySerProGlyArgAspGlySerProGly 1028
Qy 1417 ATGCAAGATGCCAGATCTGATACCATAGAGCCCGGAGATTCTCTTGGATCCAG 1476
Db 1029 AlalysGlyAspArgGlyGluThrGlyProAlaGlyProGlyAlaProGlyAlaPro 1048
Qy 1477 TTTCCAGCAAAACCCACACCCAGCTC-----ACACAGCAAAACAAATGAGAGGCCCA 1530
Db 1049 ValAlaProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluThrGlyPro 1068
Qy 1531 GAGCCTGAAGCAACAGTGTCCCTTCTGGCTGTGTGGAGCC 1572
Db 1069 AlaglyProAlaGly-----ProValGlyProValGlyAla 1080

RESULT 14
US-08-642-255-73
; Sequence 73, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-73

Alignment Scores:
Pred. No.: 7.92e-07
Score: 176.50
Length: 633
Matches: 160
Percent Similarity: 29.25%
Conservative: 26
Best Local Similarity: 25.16%
Mismatch: 245
Query Match: 5.64%
Indels: 205
DB: 1
Gaps: 36

US-09-762-021a-1 (1-1710) x US-08-642-255-73 (1-633)
Qy 25 GCAGAAGCTCTGGAGGAAGCTGGAGCAAGACCTCGACTTTGGAGGCTTCAGCCAGG 84
Db 89 AlaglnGlyProAlaGlyProGlyGlySerArgGlyAspProGlyProGlyAlaGln 108
Qy 85 CCAGGACAGATGAGGGGGCTGTCTATGAAAGGCCCTCCCTATGAGCA----- 135
Db 109 GlyProAlaGlyProGlyGlySerArgGlyAspProGlyProGlyAlaGlnGlyPro 128
Qy 136 -----GGCAGCTATCTGGAGCGGGGATCCCTCCAGACAGCCCA----- 177
Db 129 AlaglyProGlyGlySerArgGlyAspProGlyProGlyAlaGlnGlyProAlaGly 148
Qy 178 -----CCAGAGGACCTAGAGACACAGCTCCACCATCCCAAG 216
Db 149 ProGlyGlySerArgGlyAspProGlyProGlyAlaHisGlyProAlaGlyProLys 168
Qy 217 GCCCTGCCAGCCACACAGTGTCCCGAGAACCAAGTCTTACTGTCTCTCCCAAG 276
Db 169 GlyAlaHisGlyPro-----AlaGlyProLys 177
Qy 277 GCGTCTCTTCCCGAGGAGGAGGAGGAGGAGGAGTGTGAACCATGTCTTAAG 336
Db 178 GlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAlaGly-----ProLys 195
Qy 337 GGACATTGAGCTGTTCATGGAAAGCTGGAGAGGAGGAGGAGGAGGAGGAGGAGAA 396
Db 196 Gly-----AlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGlyPro 212
Qy 397 GAAATTTGGGAAAAAACAAGGAGGAGG-----AGGTCTCACCCAGGACAGTACAT 450
Db 213 ProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGly----- 229
Qy 451 TGACTGCTCCAGAGATCAAGTACAGCTTCAACCTCTGGGAGGCTGCCACCTG--- 507
Db 230 -----ProGlyAlaGlnGlyProAlaGly 238
Qy 508 -----GCTGAAGGAGACAAGTGCCCTGAGCTCGTACATCTCTTCAAGTCCCT 558
Db 239 ProGlyGlySerArgGlyAspProGlyPro----- 248
Qy 559 GAATTTATCTTGGC---CAGGTGCCCTGAGGTGGCTAGCAGCCCAAGTATCTCACC 615
Db 249 -----ProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGly 265
Qy 616 CCTCCTCACCCCTAAAGCTATCAACCTGTCTAGTCTGTCTAGGCCACCTGAGAGTAA 675
Db 266 ProGly-----AlaGlnGlyProAla--GlyPro----- 275
Qy 676 CTTTGGATGGGTGGGCCCCAGCTGAGCA-----CTAGCCGGGCCGA 720
Db 276 -----GlyGlySerArgGlyAspProGlyProGlyAlaGlnGlyProAlaGlyProG 294
Qy 721 CTGACAGGCGATGAGCCCTGCGCTACCAACCCACATCTCAGATGACTGCAACTTC 780
Db 294 lyGlySerArgGlyAspProGlyProGlyAlaHisGly----- 307
Qy 781 AGACCCCTCCAGCAAGCACCTTAGATACAGGACCTGTTTCCCTTCGGGGGGAAG 840
Db 308 -----ProAla--GlyProLysGlyAlaHisGlyProAlaGlyProLys----- 321
Qy 841 TCATAGTTAGGAGCACCTCACACTTCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 322 -----GlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAl 336
Qy 901 G-----CCTGGGGACCCC-----AACTCCAGGCC 924
Db 336 aglyProLysGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAsp-ProGlyP 356
Qy 925 CTCAGCCCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTGTACAGTTGAAGC 984
Db 356 roProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGly-----Asp 372


```
Db 239 roAlaGlyProProGlyAlaProGlyProAlaGlyPro----- 251
Qy 778 TCCAGAGCCCTCCAGCCCAAGCACCCTTAGGATACCAGGACCCTGTTCCCTTCGCGCGGG 837
Db 252 -----ProGlyAlaProGlyPro----- 257
Qy 838 AAGTCATAGGTTAGGAGCACCCTCACACTTTCCTCAGGAGAAGACACACAACCATGACCC 897
Db 258 -----AlaGlyProPro----- 261
Qy 898 TCAGCCTGGGGACCCCAACTCCA---GGCCCTCCAGCCCCCAACCTGCCAGCCAGCCCT 954
Db 262 -----GlyAlaProGlyProAlaGlyProProGlyAlaHis-GlyProAlaGlyPro 278
Qy 955 GAAAATGCAAGCTCTGTACGAGTTTGAAGCTAGGAAACCCAGCGGAACCTGACTGTGGTCCA 1014
Db 279 LysGlyAlaHis-----GlyPro 284
Qy 1015 GGGAGAGAAGCTGGAGGTTCTGGACCACAGCAGCGGTGGTGGCTGGAAGAATGAGGC 1074
Db 285 AlaGlyProLysGlyAlaHisGlyPro-----Ala 294
Qy 1075 GGGAGCGAGCGGTACATTCACGAACAATCCT---GGAGCCCT---ACAGCCGGGAC 1128
Db 295 GlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaProGlyProAlaGlyPro 314
Qy 1129 CCCTGGGACCAGCGCCAGTCACCTCTCGGTTCCAAATGCTTCGACTTAGCTCGAGGCC 1188
Db 315 ProGlyAlaProGlyProAlaGlyProProGlyAlaProGly----- 328
Qy 1189 TGAAGAGGTCACAGACTGGCTGCAGGAGAGAACTTCTCCACTGCCACGGTGAGGACACT 1248
Db 329 -----ProAlaGly----- 331
Qy 1249 TGGTCCCTGACGGGAGCCACTTTCGCATAAGACTGGGAGCTACAGATGCTATG 1308
Db 332 -----ProProGlyAlaProGlyProAla-----GlyProProGlyAlaPro 345
Qy 1309 TCCACAGGAGCCCCAGCAAT---CCTGTCCCGCTGA-----GGCTGCAGAAGAT 1359
Db 346 GlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyPro 365
Qy 1360 GCTGGGATAAGCCTTAGGCACCAGCTTAGACACCTCCAAGAACCAGGCCCGCTCATG 1419
Db 366 AlaGly-----ProProGlyAlaProGlyProAlaGlyProProGlyAlaPro----- 381
Qy 1420 CAAGATGGCAGATGTATCCATTAGAGCCCGAGAATTCTCTCTGGATCCAGTTT 1479
Db 382 -----GlyProAlaGlyProProGly 388
Qy 1480 GCAGCAAAACCCACCCAGCTCACACAGCAAAACAAATGGACAGGCCCGAGGCTGAA 1539
Db 389 AlaProGlyProAlaGlyPro-----ProGlyAlaPro 399
Qy 1540 GCAACACAGTGTCCCTTCTGGCTGTGTGGA---GCCCTCCCAAGTAACCACTATTATTT 1596
Db 400 GlyProAlaGlyProProGlyAlaProGlyProAlaGlyPro----- 413
Qy 1597 TACCTCTTCCCAACCTGGAGCATTTATGCTAGGCTTGTCAAGAATCTGTTCAAGTCCC 1656
Db 414 -----ProGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyPro 428
Qy 1657 TCTCCTTCTCAATAAAGCAT 1677
Db 429 AlaGlyProLysGlyAlaHis 435
```

Search completed: February 25, 2003, 19:45:22
Job time : 42.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 05:25:51 ; Search time 3042 Seconds
(without alignments)
16359.565 Million cell updates/sec

Title: US-09-762-021A-1

Perfect score: 1710

Sequence: 1 ggcagagcactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_sts.*

12: gb_sy.*

13: gb_un.*

14: gb_vi.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_sts.*

28: em_un.*

29: em_vi.*

30: em_htg_hum.*

31: em_htg_inv.*

32: em_htg_other.*

33: em_htg_mus.*

34: em_htg_pin.*

35: em_htg_rod.*

36: em_htg_man.*

37: em_htg_vrt.*

38: em_sy.*

39: em_htgo_hum.*

40: em_htgo_mus.*

41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1696.6	99.2	2220	6	AX058186 Sequence
2	1683.6	98.5	2255	9	BC012926 Homo sapi
3	1508.6	88.2	2148	9	AK025175 Homo sapi
4	1424.4	83.3	1973	9	AY074930 Homo sapi
5	1334.6	78.0	2159	6	AX322735 Sequence
6	918.2	53.7	2275	10	BC014734 Mus muscu
7	913.2	53.4	2329	10	AY074932 Mus muscu
8	389	22.7	401	6	AX407643 Sequence
9	324.4	19.0	133451	9	AL158847 Human DNA
10	322.8	18.9	162480	2	AC092860 Pan trogl
11	218.4	12.8	224	11	G25672 human STS E
12	218.4	12.8	224	11	G41866 SHGC-32779
13	155.6	9.1	197	6	AR044148 Sequence
14	117.6	6.9	113285	10	AC018461 Mus muscu
15	117.6	6.9	198459	2	AC079042 Mus muscu
16	110.8	6.5	185397	2	AC095838 Rattus no
17	102.6	6.0	185397	2	AC095838 Rattus no
18	91.8	5.4	2527	9	AY074929 Homo sapi
19	91.8	5.4	2646	9	BC002474 Homo sapi
20	91.8	5.4	3074	9	AK025824 Homo sapi
21	91.8	5.4	3102	9	AK025588 Homo sapi
22	88	5.1	2557	9	AY074928 Homo sapi
23	86.4	5.1	2201	9	BC015763 Homo sapi
24	86.4	5.1	2228	9	AK000265 Homo sapi
25	86.4	5.1	2685	6	AX399900 Sequence
26	85	5.0	1829	9	AF282167 Homo sapi
27	84.8	5.0	2169	9	BC004907 Homo sapi
28	83.2	4.9	125020	9	AF429315 Homo sapi
29	79.8	4.7	3079	10	BC005492 Mus muscu
30	70.4	4.1	2151	10	AY074931 Mus muscu
31	70.4	4.1	2995	10	BC027043 Mus muscu
32	66	3.9	2233	9	AK094539 Homo sapi
33	64.2	3.8	125020	9	AF429315 Homo sapi
34	63.8	3.7	1827	3	AY075572 Drosophil
35	63.8	3.7	3035	3	AF208262 Drosophil
36	62.6	3.7	90816	2	AC017493 Drosophil
37	62.6	3.7	102389	3	AC004546 Drosophil
38	62.6	3.7	165652	3	AC092222 Drosophil
39	62.6	3.7	299378	3	AE003588 Drosophil
40	56.8	3.3	1977	9	AK057052 Homo sapi
41	56.4	3.3	1947	9	AK027765 Homo sapi
42	56.2	3.3	3365	10	BC009098 Mus muscu
43	55	3.2	514	6	AX321026 Sequence
44	55	3.2	3832	9	HSU12535 Human epide
45	55	3.2	3892	9	BC030010 Homo sapi

ALIGNMENTS

RESULT 1

AX058186

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

REFERENCE

AUTHORS

TITLE

AX058186 Sequence 56 from Patent WO0077040. 2220 bp DNA linear PAT 17-JAN-2001

AX058186.1 GI:12310687

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2220)

Yue.H., Tang.Y.T., Hillman,J.L., Lal,P., Bandman,O., Baughn,M.R.,

Azimzai,Y., Yang,J., Reddy,R. and Lu,D.A.

Human intracellular signaling molecules

JOURNAL Patent: WO 0077040-A 56 21-DEC-2000;

Incyte Genomics, Inc. (US)

FEATURES Location/Qualifiers

Source

1..2220

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 549 a 696 c 593 g 382 t

ORIGIN

Query Match 99.2%; Score 1696.6; DB 6; Length 2220;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1699; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAGAGCACTGAAGACAGCCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAAAGACC 60

Db 518 GCGAGAGCACTGAAGACAGCCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAAAGACC 577

Qy 61 TCGACTTGGAGGCTTACAGCAGGCGCAGGACAGATGGAGGGGCTGTATGGAAAGGCC 120

Db 578 TCGACTTGGAGGCTTACAGCAGGCGCAGGACAGATGGAGGGGCTGTATGGAAAGGCC 637

Qy 121 GCTCCCTATGAGGAGGACGACCTATCTGAGCGCGGGATCCCTCCAGAACAGCCCAACA 180

Db 638 GCTCCCTATGAGGAGGACGACCTATCTGAGCGCGGGATCCCTCCAGAACAGCCCAACA 697

Qy 181 GAGGACCTTAGAGCACAGCCTCCACCATCCCAAGGCCCTTGCACGCCACACCAAGTGC 240

Db 698 GAGGACCTTAGAGCACAGCCTCCACCATCCCAAGGCCCTTGCACGCCACACCAAGTGC 757

Qy 241 CCGAACAACAGTGCCTTACTCTGCTCCTCCAAAGGGCGTCTTCCCGCGAGGAGCC 300

Db 758 CCGAACAACAGTGCCTTACTCTGCTCCTCCAAAGGGCGTCTTCCCGCGAGGAGCC 817

Qy 301 AGAGAGGACGAGGAAGTGTGAACCATGTCTTAAGGACATTTAGCTGTTCATGGGAAA 360

Db 818 AGAGAGGACGAGGAAGTGTGAACCATGTCTTAAGGACATTTAGCTGTTCATGGGAAA 877

Qy 361 GCTGGAGAAGGCCCAAGCAAGACAGCAGGAAGAAGAAATTTGGGAAAAAACAAGGA 420

Db 878 GCTGGAGAAGGCCCAAGCAAGACAGCAGGAAGAAGAAATTTGGGAAAAAACAAGGA 937

Qy 421 CCAGGAGGTCTACCCAGGACAGTACATTTGCTGCTCCAGAGATCAAGTACAGTT 480

Db 938 CCAGGAGGTCTACCCAGGACAGTACATTTGCTGCTCCAGAGATCAAGTACAGTT 997

Qy 481 CAACCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAAGTCCCTGAGCTCGTACA 540

Db 998 CAACCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAAGTCCCTGAGCTCGTACA 1057

Qy 541 CATCCTCTCAAGTCCCTGAACCTTCATCTGCGCCAGGTGCCCTGAGGCTGGCCCTAGCAGC 600

Db 1058 CATCCTCTCAAGTCCCTGAACCTTCATCTGCGCCAGGTGCCCTGAGGCTGGCCCTAGCAGC 1117

Qy 601 CCAAGTATCTACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTAAG 660

Db 1118 CCAAGTATCTACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTAAG 1177

Qy 661 CCACCTCTGAGAGTAACTTTGGATGGGTGGGCCAGCCTGGACCACTAGCCGGGCCGA 720

Db 1178 CCACCTCTGAGAGTAACTTTGGATGGGTGGGCCAGCCTGGACCACTAGCCGGGCCGA 1237

Qy 721 CTGGACAGCGATGAGCCCTGCCCTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 780

Db 1238 CTGGACAGCGATGAGCCCTGCCCTACCAACCCACATTTCTCGGATGACTGGCAACTTCC 1297

Qy 781 AGAGCCCTCCAGCAAGCAACCTTTAGGATACAGGACCCCTGTTCCTTCCGCGGGGAAG 840

Db 1298 AGAGCCCTCCAGCAAGCAACCTTTAGGATACAGGACCCCTGTTCCTTCCGCGGGGAAG 1357

Qy 841 TCATAGTTAGGAGCACTCACATTTCTCAGGAGAGACACACACCACTGACCTCA 900

Db 1358 TCATAGTTAGGAGCACTCACATTTCTCAGGAGAGACACACACCACTGACCTCA 1417

Qy 901 GCCTGGGAGCCCAACATCCAGGCCCTCCAGCCCCCAAACTCCCGCAGCAGCCCTGAAAAT 960

Db 1418 GCCTGGGAGCCCAACATCCAGGCCCTCCAGCCCCCAAACTCCCGCAGCAGCCCTGAAAAT 1477

Qy 961 GCAAGTCTTTACGAGTTTGAAGCTAGGAACCCAGCGGAACCTGACTGTGGTCCAGGGAGA 1020

Db 1478 GCAAGTCTTTACGAGTTTGAAGCTAGGAACCCAGCGGAACCTGACTGTGGTCCAGGGAGA 1537

Qy 1021 GAAGCTGGAGGTTCTGGACACACAGCAAGCGGTGGTGGCTGAGTGAAGATGAGGCGGGACG 1080

Db 1538 GAAGCTGGAGGTTCTGGACACACAGCAAGCGGTGGTGGCTGAGTGAAGATGAGGCGGGACG 1597

Qy 1081 GAGCGGCTACATTTCCAAAGCAACATCTCTGGAGCCCTACAGCGGGGAGCCCTGGGAGCCCA 1140

Db 1598 GAGCGGCTACATTTCCAAAGCAACATCTCTGGAGCCCTACAGCGGGGAGCCCTGGGAGCCCA 1657

Qy 1141 GGGCAGTCAACCTCTCGGGTTCCAATGCTTTCGACTTTCGAGCTTCGAGGCTTCGAGAGGTCA 1200

Db 1658 GGGCAGTCAACCTCTCGGGTTCCAATGCTTTCGACTTTCGAGCTTCGAGGCTTCGAGAGGTCA 1717

Qy 1201 AGACTGGCTGAGGACAGAGAACTTCTCCACTGCCACGCTGAGGACACTTGGGTCCCTGAC 1260

Db 1718 AGACTGGCTGAGGACAGAGAACTTCTCCACTGCCACGCTGAGGACACTTGGGTCCCTGAC 1777

Qy 1261 GGGAGGCGAGTACTTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCACAGGAGGC 1320

Db 1778 GGGAGGCGAGTACTTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCACAGGAGGC 1837

Qy 1321 CCCAGCAATCTGTCGGGCTGGAGGCTGTCAGAGGATGCTGGGGATAAGCCCTTAGGC 1380

Db 1838 CCCAGCAATCTGTCGGGCTGGAGGCTGTCAGAGGATGCTGGGGATAAGCCCTTAGGC 1897

Qy 1381 ACCAGCTTAGACACCTCCAAAGAACAGGCCCGCTGATGCAAGATGGCAGATCTGATACC 1440

Db 1898 ACCAGCTTAGACACCTCCAAAGAACAGGCCCGCTGATGCAAGATGGCAGATCTGATACC 1957

Qy 1441 CATTAGAGCCCGAGAAATTCCTCTCTGGATCCCAAGTTTGCAGCAAAACCCACACCCAG 1500

Db 1958 CATTAGAGCCCGAGAAATTCCTCTCTGGATCCCAAGTTTGCAGCAAAACCCACACCCAG 2017

Qy 1501 CTCACAGCAAAACAAATGGACAGGCCAGGCTGAAGCAACAGTGTCCCTTCTGCG 1560

Db 2018 CTCACAGCAAAACAAATGGACAGGCCAGGCTGAAGCAACAGTGTCCCTTCTGCG 2077

Qy 1561 TGTGTGGAGGCTCCCAAGTAAACCACTTATTTATTTTACCTCTTTCCCAACCTGGAGCA 1620

Db 2078 TGTGTGGAGGCTCCCAAGTAAACCACTTATTTATTTTACCTCTTTCCCAACCTGGAGCA 2137

Qy 1621 TTTATGCTAGGCTGTCAAGAACTGTGTCAGTCCCTCTCCTTCTCAATAAAGCATCTT 1680

Db 2138 TTTATGCTAGGCTGTCAAGAACTGTGTCAGTCCCTCTCCTTCTCAATAAAGCATCTT 2197

Qy 1681 CAAGCTTCTAAAAAATAA 1703

Db 2198 CAAGCTTCTAAAAAATAA 2220

RESULT 2

BC012926

LOCUS

DEFINITION

Homo sapiens, Similar to hypothetical protein FLJ21522, clone

MGC:16817 IMAGE:3853503, mRNA, complete cds.

ACCESSION

BC012926

VERSION

BC012926.1 GI:15277845

KEYWORDS

MGC.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 2255)

AUTHORS

Straussberg, R.

TITLE

Direct Submission

JOURNAL

Submitted (20-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgabps@mail.nih.gov

Tissue procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center

Center code: BCM-HGSC

Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>

Contact: villalonga@bcm.tmc.edu

Villalon, D.K., Luna, R.A., Haile, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 13 Row: a Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA qi: 10437637.

FEATURES
source

```

1. .2255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:16817 IMAGE:
/tissue_types="Colon, ade
/clone_lib="NIH_MGC_65"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT
146. .1930

```

146 1930

```

/codon_start=1
/product="Similar to hypothetical protein FLJ21522"
/protein_id="AAH12926.1"
/db_xref="GI:15277846"
/translation="MSRSPRAIYLHRKEYSONLTSEPTLQHRVHMLMTCOGSRV
QGPEDAKLIFENDAGRWYMSQDILQVRDGLWQLDITKEEDSLRJDLSIQAMVA
LNTCSNISLTIVPEGPLGTSTLLFCQVGAERLSTLQKALEEELSPGRLG
LQSGDRWRGPAWMLPMEQAYRGLPPEQPHORTLHSLPSPRPILPRTSRAE
PSATLLPPRSSSPDEPDEVLNVLURDIELPMGLKKAQKATSRKKKFKGKNN
QGGUQTQYIDCFQKTIKYFNGLGRUATLWKESTSAPELVIYLFKSLILARCPAEG
AAQVISPFLPTAKILNQSLSPSNLWMLGPAWTTSRADTGDLEPLQPTFFSD
WQLPEPSSQAPLQGVDPVSLRGSHRLGSTSHFPOEKTHNDHPDGGPNRSPSPKPA
QPALKMQVLYVEARPNRELTYYQGEKLEVDLHCKRWMLVKNKAGRGYTPSNILEP
QPQTGTPQSGSRVPMRLSRPSEVTDWQAENFSTATVYITGSLGSQLLRIRP
ELQMLCPQEAAPRILSRLEAVRMGLTSP"
705 c 607 c 386 t

```

BASE COUNT
ORIGIN

Query Match 98.5%; Score 1683.6; DB 9; Length 2255;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1699; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

Qy	1	GGCAGCGGACTGAAGACCAAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAA---AG	57
Db	550	GGCAGAGCGGACTGAAGACCAAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAAAGCAG	609
Qy	58	ACCTCGACTTTGGAGGCCCTTCAGCCAGGCCACAGGACAGATGGAGGGGGGCTGCTATGGAAG	117
Db	610	ACCTCGACTTTGGAGGCCCTTCAGCCAGGCCACAGGACAGATGGAGGGGGCTGCTATGGAAG	669
Qy	118	GCGCCTCCCTATGGAGCAGGACAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCA	177
Db	670	GCGCCTCCCTATGGAGCAGGACAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCA	729
Qy	178	CCAGAGAGACCTTAGAGCAGACGCTCCACCATCCCAAGGCCCTGCCAGGCCACACAG	237
Db	730	CCAGAGAGACCTTAGAGCAGACGCTCCACCATCCCAAGGCCCTGCCAGGCCACACAG	789
Qy	238	TGCCCGAGAACCAAGTGCCCTTTACTCTGCCTCTCTCCAAAGGCGTCTCTTCTCCCGCAGGA	297

Db	1362	-----	-----	-----	1361
QY	901	GCCTGGGACCCCAACTCCAGGCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAT	960		
Db	1362	-----	-----	-----	1401
QY	961	GCAAGTCTTGTACGAGTTTCAAGCTAGGAACCCACGGGAACCTGACTGTGTCTCCAGGAGA	1020		
Db	1402	GCAAGTCTTGTACGAGTTTCAAGCTAGGAACCCACGGGAACCTGACTGTGTCTCCAGGAGA	1461		
QY	1021	GAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGC	1080		
Db	1462	GAACTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGC	1521		
QY	1081	GAGCGCTACATTCACAGCAACATCTGGAGCCCTACAGCCGGGACCCCTGGGACCCA	1140		
Db	1522	GAGCGCTACATTCACAGCAACATCTGGAGCCCTACAGCCGGGACCCCTGGGACCCA	1581		
QY	1141	GGGCCAGTCAACCTCTCGGTTTCCAATGCTTTCGACTTAGCTCGAGGCTGAAGAGTCA	1200		
Db	1582	GGGCCAGTCAACCTCTCGGTTTCCAATGCTTTCGACTTAGCTCGAGGCTGAAGAGTCA	1641		
QY	1201	AGACTGGCTGAGGACAGAACTTCTCCACTGCCCGGTGAGGACACTTGGGTCCCTGAC	1260		
Db	1642	AGACTGGCTGAGGACAGAACTTCTCCACTGCCCGGTGAGGACACTTGGGTCCCTGAC	1701		
QY	1261	GGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGC	1320		
Db	1702	GGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGC	1761		
QY	1321	CCCAGAACTCTCTCCGGTGGAGGCTGTGCAAGAGTGTGGGATAAGCCCTTAGGC	1380		
Db	1762	CCCAGAACTCTCTCCGGTGGAGGCTGTGCAAGAGTGTGGGATAAGCCCTTAGGC	1821		
QY	1381	ACCAGCTTAGACACCTCCAAAGAACAGGCCCTGATGCAAGATGCGAGATCTGATACC	1440		
Db	1822	ACCAGCTTAGACACCTCCAAAGAACAGGCCCTGATGCAAGATGCGAGATCTGATACC	1881		
QY	1441	CATTAGAGCCCGAGAACTCTCTCTGGATCCAGTGTGCAGCAACCCACACCCAG	1500		
Db	1882	CATTAGAGCCCGAGAACTCTCTCTGGATCCAGTGTGCAGCAACCCACACCCAG	1941		
QY	1501	CTCACAGCAAAACAATGACAGGCCAGAGGCTGAAGCAACAGTGTCCCTTCTGGC	1560		
Db	1942	CTCACAGCAAAACAATGACAGGCCAGAGGCTGAAGCAACAGTGTCCCTTCTGGC	2001		
QY	1561	TGTGTTGGAGCCTCCCGAGTAACCACTATTATTTTACCTCTTTCCCAACCTGGAGCA	1620		
Db	2002	TGTGTTGGAGCCTCCCGAGTAACCACTATTATTTTACCTCTTTCCCAACCTGGAGCA	2061		
QY	1621	TTTATGCTAGGCTTGTCAAGAATCTGTTTCAGTCCCTCTCTCTCAATAAAGCATCTT	1680		
Db	2062	TTTATGCTAGGCTTGTCAAGAATCTGTTTCAGTCCCTCTCTCTCAATAAAGCATCTT	2121		
QY	1681	CAAGCTTGTAAAAAATAAANGA	1706		
Db	2122	CAAGCTTGTAAAAAATAAANGA	2147		
RESULT 4					
AY074930					
LOCUS					
DEFINITION					
Homo sapiens epidermal growth factor receptor pathway substrate 8					
related protein 3 (EPS8R3) mRNA, complete cds.					
ACCESSION					
AY074930					
VERSION					
AY074930.1 GI:18655334					
KEYWORDS					
SOURCE					
ORGANISM					
Homo sapiens.					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE					
1 (bases 1 to 1973)					

AUTHORS	Scita,G., Confalonieri,S., Offenhauser,N., Borgonovo,A., Tocchetti,A., Romano,P. and Di Fiore,P.P.				
TITLE	Cloning and characterization of novel members of the Eps8 protein family				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 1973)				
AUTHORS	Scita,G., Confalonieri,S., Offenhauser,N., Borgonovo,A., Tocchetti,A., Romano,P. and Di Fiore,P.P.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JAN-2002) Experimental Oncology, European Institute of Oncology, Via Ripamonti 435, Milano 20141, Italy				
FEATURES	Location/Qualifiers				
source	1..1973				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
gene	1..1973				
	/gene="EPS8R3"				
CDS	112..1893				
	/gene="EPS8R3"				
	/note="EPS8 related protein 3"				
	/codon_start=1				
	/product="epidermal growth factor receptor pathway substrate 8 related protein 3"				
	/protein_id="AAL76119.1"				
	/db_xref="GI:18655335"				
	/translation="MSRPSRAIYLHRKEYSONLTSEPTLQHRVHEHLMTCQGSORV OGPEDALOKLFEMDAQGWSODLILQVRDGLQLLDLETKEELSYRLDSIQANMVA LNTCSYNSILSITVOEPGLPGTSTLLFQCEVGAERLXTSLQKALEEELERPLRGLGL QSPQWRGPMARPELPEGPPIPPPEPHQRTLEHSLPPSPRPHRTSAREP SFTLPPRRSSPEDDEVLNHLRDIELFMKLEAKAQRKRRKFKKRRDQ GGTQAQYIDCFQIKHSFNLLGLRLATWLKTSAPLNFILFKSLNLFARCEAGLA AOVISPLLTAKAINLLQSCSPESNLWMLGPAWTTSRADWTGDEPIPYOPTFSDDW OLPESSOAPILGYDPVSLRSGHSLGTSHPQKTHNHDPQDNPNSRSPKPAQ PALMOLVLEFEARNPRELTYYOGKEVLVLDHSEKMLVKNKNEAGSGVYIPSNILPELO PGTPTGQSPSRVPMRLSRSPPEVTWMLQENFSTATVTRLGSLTSGSOLLIRPGE LOMLCPQEPAPRLSLREAVRMLGTSIP"				
BASE COUNT	481 a	620 c	549 g	321 t	2 others
ORIGIN					
Query Match	83.3%; Score 1424.4; DB 9; Length 1973;				
Best Local Similarity	99.3%; Pred. No. 0;				
Matches 1450;	Conservative 0; Mismatches 8; Indels 2; Gaps 2;				
QY	1	GGCAGAGCGACTGAACACAGCGCTGCAGAGGCTCTGGAGGAAGAGCTGCAGCAAGACC	60		
Db	516	GGCAGAGCGACTGAACACAGCGCTGCAGAGGCTCTGGAGGAAGAGCTGCAGCAAGACC	575		
QY	61	TGACCTTGGAGGCTTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGTATGAAAGGCC	120		
Db	576	TGACCTTGGAGGCTTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGTATGAAAGGCC	635		
QY	121	GCTCCCTATGGAGCGGCGCTATCTGGAGCGGGGATCCCTCCAGACAGCCGCCACCA	180		
Db	636	GCTCCCTATGGAGCGGCGCTATCTGGAGCGGGGATCCCTCCAGACAGCCGCCACCA	695		
QY	181	GAGGACCTTAGAGCAGCAGCTCCACCATCCCAAGCGCCCTGCCACGCCACACAGTGC	240		
Db	696	GAGGACCTTAGAGCAGCAGCTCCACCATCCCAAGCGCCCTGCCACGCCACACAGTGC	755		
QY	241	CCGAGAACCAAGTGCCTTTACTCTGCCTCTCTCAAGCGGCTCTCTTCCCGCCGAGGCC	300		
Db	756	CCGAGAACCAAGTGCCTTTACTCTGCCTCTCTCAAGCGGCTCTCTTCCCGCCGAGGCC	815		
QY	301	AGAGAGGAGGAGGAGTCTGTAACCATCTCTTAAGGGACATGAGCTGTTCATGGGAA	360		
Db	816	AGAGAGGAGGAGGAGTCTGTAACCATCTCTTAAGGGACATGAGCTGTTCATGGGAA	875		
QY	361	GCTGGAGAGAGGCGCCAGGCAAGACACAGAGGAAGAATAATTTGGGAAAAAACAGGA	420		
Db	876	GCTGGAGAGAGGCGCCAGGCAAGACACAGAGGAAGAATAATTTGGGAAAAAACAGGA	935		
QY	421	CCAGGAGGAGTCTTACCCAGGACAGTACATGACTGCTTCCAGAGATCAAGTACAGCTT	480		

Db 936 CCAGGGAGTCTACCCAGGACAGTACATTTGACTGCTTCCAGAAAGATCAAGCACAGCTT 995
Qy 481 CAACCTCTCTGGAAGGTGGCCACCTGGCTGAAGGAGACAAGTCCCTCGAGCTCGTACA 540
Db 996 CAACCTCTCTGGAAGGTGGCCACCTGGCTGAAGGAGACAAGTCCCTCGAGCTCGTACA 1055
Qy 541 CATCTCTTCAAGTCCCTGAATTCATCTCTGGCCAGGTGGCCCTGAGGCTGAGCCTAGCAGC 600
Db 1056 CATCTCTTCAAGTCCCTGAATTCATCTCTGGCCAGGTGGCCCTGAGGCTGAGCCTAGCAGC 1115
Qy 601 CCAAGTATCTACCCCTCTCCACCCCTTAAGCTATCAACCTGTACAGTCTCTCTTAAG 660
Db 1116 CCAAGTATCTACCCCTCTCCACCCCTTAAGCTATCAACCTGTACAGTCTCTCTTAAG 1175
Qy 661 CCCACCTTGAGAGTAACCTTTTGATGGGCTTGGGCCAGCCTGGACCACTAGCCGGCGGA 720
Db 1176 CCCACCTTGAGAGTAACCTTTTGATGGGCTTGGGCCAGCCTGGACCACTAGCCGGCGGA 1235
Qy 721 CTGGACAGGGGATGAGCCCTTGCCTTACCAACCCACATTTCTAGATGACTGGCAACTTCC 780
Db 1236 CTGGACAGGGGATGAGCCCTTGCCTTACCAACCCACATTTCTGGATGACTGGCAACTTCC 1295
Qy 781 AGAGCCCTCCAGCCCAAGCACCTTTAGGATACCGAGACCCCTGTTTCCCTTCCGCCGGGAAG 840
Db 1296 AGAGCCCTCCAGCCCAAGCACCTTTAGGATACCGAGACCCCTGTTTCCCTTCCGCCGGGAAG 1355
Qy 841 TCATAGGTTAGGAGCACCTCACACTTTCCTCAGGAGAAGACACACCAACATGACCTCA 900
Db 1356 TCATAGGTTAGGAGCACCTCACACTTTCCTCAGGAGAAGACACACCAACATGACCTCA 1415
Qy 901 GCCTGGGACCCCAACTCCAGGCGCTCCAGCCCAACCTCCAGGAGGAGCCCTGAAAT 960
Db 1416 GCCTGGGACCCCAACTCCAGGCGCTCCAGCCCAACCTCCAGGAGGAGCCCTGAAAT 1475
Qy 961 GCAAGCTTTGAGAGTTGAGCTAGGAACCCAGGGAACCTGACTGTGCTCAGGAGA 1020
Db 1476 GCAAGCTTTGAGAGTTGAGCTAGGAACCCAGGGAACCTGACTGTGCTCAGGAGA 1535
Qy 1021 GAAGCTGGAGGTTCTGGACACACAGCAGCGTGTGCTGAGGAATGAGCGGAGC 1080
Db 1536 GAAGCTGGAGGTTCTGGACACACAGCAGCGTGTGCTGAGGAATGAGCGGAGC 1595
Qy 1081 GAGCGGCTTACATTCGAAGCAACATCTCTGGAGCCCTACAGCCGGGAGCCCTGGGACCCA 1140
Db 1596 GAGCGGCTTACATTCGAAGCAACATCTCTGGAGCCCTACAGCCGGGAGCCCTGGGACCCA 1655
Qy 1141 GGGCCAGTCAACCTCTCGGCTTCCAATGCTTCGACTTACGCTGAGGCTGAAGAGGTAC 1200
Db 1656 GGGCCAGTCAACCTCTCGGCTTCCAATGCTTCGACTTACGCTGAGGCTGAAGAGGTAC 1715
Qy 1201 AGACTGGCTGAGGAGAGAACTTCTCCACTGCCACGCTGAGGACACTTGGGTCCCTGAC 1260
Db 1716 AGACTGGCTGAGGAGAGAACTTCTCCACTGCCACGCTGAGGACACTTGGGTCCCTGAC 1775
Qy 1261 GGGGAGCAGCTACTTCGCATTAAGACCTTGGGAGCTACAGATGCTATGTCCACAGGAGGC 1320
Db 1776 GGGGAGCAGCTACTTCGCATTAAGACCTTGGGAGCTACAGATGCTATGTCCACAGGAGGC 1835
Qy 1321 CCCAGCAATCTGTCCCGGCTGGAGGCTGTCAGAAAGATGCTGGGGATAGCCCTTAGGC 1380
Db 1836 CCCAGCAATCTGTCCCGGCTGGAGGCTGTCAGAAAGATGCTGGGGATAGCCCTTAGGC 1894
Qy 1381 ACCAGCTTAGACACTTCCAAGAACAGCCCGCTGATGCAAGATGGCAGATCTGATACC 1440
Db 1895 ACCAGCTTAGACACTTCCAAGAACAGCCCGCTGATGCAAGATGGCAGATTTGATACC 1933
Qy 1441 CATTAGAGCCCGGAGAAATTC 1460
Db 1954 CATTANAGCCCGGAGAAATTC 1973

RESULT 5
AX322735

LOCUS AX322735 2159 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 10 from Patent WO0192528.
ACCESSION AX322735
VERSION AX322735.1 GI:18093725
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE 1
AUTHORS Macina, R. A.; Chen, S. Y.; Pluta, J.; Sun, Y. and Recipon, H.
TITLE Method of diagnosing, monitoring, staging, imaging and treating
colon cancer
JOURNAL Patent: WO 0192528-A 10 06-DEC-2001;
diadexus, Inc. (US)
FEATURES
Location/Qualifiers
source 1..2159
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 528 a 673 c 581 g 377 t
ORIGIN
Query Match 78.0%; Score 1334.6; DB 6; Length 2159;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1580; Conservative 0; Mismatches 29; Indels 108; Gaps 12;
Qy 1 GGCAGAGCGACTGAAGACCAAGCCCTGCAGAGGCTCTCGAGAGAGAGCTGGAGCAAGACC 60
Db 532 GGCAGAGCGACTGAAGACCAAGCCCTGCAGAGGCTCTCGAGAGAGAGCTGGAGCAAGACC 591
Qy 61 TCGACTTTGGAGGCGCTTCAGCCAGGCCAGGACAGATGAGAGGGGCGCTGCTATGGAAGGCC 120
Db 592 TCGACTTTGGAGGCGCTTCAGCCAGGCCAGGACAGATGAGAGGGGCGCTGCTATGGAAGGCC 651
Qy 121 GCTCCCTATGGAGCAGCAGCCCTATCTGGAGCCGGGATCCCTCCAGACAGCCCAACCA 180
Db 652 GCTCCCTATGGAGCAGCAGCCCTATCTGGAGCCGGGATCCCTCCAGACAGCCCAACCA 711
Qy 181 GAGACCCCTAGAGCACAGCCCTCCCAACCATCCCAAGGCCCTGCCAGCCACACAGCTGC 240
Db 712 GAGACCCCTAGAGCACAGCCCTCCCAACCATCCCAAGGCCCTGCCAGCCACACAGCTGC 771
Qy 241 CCGAGAACCAAGTGCCCTTTACTCTGCTTCCCAAGGGGCTCTCTTCCGCCAGGAGACC 300
Db 772 CCGAGAACCAAGTGCCCTTTACTCTGCTTCCCAAGGGGCTCTCTTCCGCCAGGAGACC 831
Qy 301 AGAGAGGAGCAGGAGTGTCTGAACCATGTCTTAAGGACATTCAGCTGTTTCATGGGAAA 360
Db 832 AGAGAGGAGCAGGAGTGTCTGAACCATGTCTTAAGGACATTCAGCTGTTTCATGGGAAA 891
Qy 361 GCTGGAGAAGGCCCCAGCAAGACAGCAGCAAGAAATTTGGGAAAAA---AACAA 417
Db 892 GCTGGAGAAGGCCCCAGCAAGACAGCAGCAAGAAATTTGGGAAAAAAGAGAACA 951
Qy 418 GGACCCAGGAGGTCTACCCAGGCACAGTACA-TTGACTGCTTCCAGAAATCAAGTACA 476
Db 952 GGACCCAGGAGGTCTACCCAGGCACAGTACA-TTGACTGCTTCCAGAAATCAAGTACA 1011
Qy 477 GCTTCAACCTCTGGAGGCTGGCCACCTGGCTGAAGGACAAAGTGCCCTTGAGCTCG 536
Db 1012 GCTTCAACCTCTGGAGGCTGGCCACCTGGCTGAAGGACAAAGTGCCCTTGAGCTCG 1071
Qy 537 TACACATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGCCCTGAGCTGGCCCTAG 596
Db 1072 TACACATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGCCCTGAGCTGGCCCTAG 1131
Qy 597 CAGCCCAAGTGATCTACCCCTCTCACCCTTAAGCTATCAACCTGCTACAGTCTCTGTC 656
Db 1132 CAGCCCAAGTGATCTACCCCTCTCACCCTTAAGCTATCAACCTGCTACAGTCTCTGTC 1191
Qy 657 TAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCCAAGCTGGACCACTAGCCGGG 716
Db 1192 TAAGCTCACCTGAGAGTAACCTTTGGATGGGTTGGGCCCAAGCTGGACCACTAGCCGGG 1251

QY	717	CCGACTGGACAGCGGATGAGCCCTGCGCTACCAACCCACATTTCTCAGATGACTGGCAAC	776
Db	1252	CCGACTGGACAGCGGATGAGCCCTGCGCTACCAACCCACATTTCTCAGATGACTGGCAAC	1311
QY	777	TTCCAGAGCCCTCCAGCCAGCACCTTGTAGATACACAGGACCCCTGTTCCCTTCGCGGG	836
Db	1312	TTCCAGAGCCCTCCAGCCAGCACCTTGTAGATACACAGGACCCCTGTTCCCTTCGCGGG	1366
QY	837	GAAGTCATAGTTAGGAGCACCTCACACTTCTCTCAGGAGAGACACACCAACCATGACC	896
Db	1367	-----	1366
QY	897	CTCAGCCTGGGACCCCAACTCCAGGCCCTTCCAGCCCAAACTGCCCAAGCCAG-CCCTTG	955
Db	1367	-----GGCCTCCAGTCCCAAACTGCCCAAGCCAGTCCCTG	1402
QY	956	AAATGCAAGTCTGTACGAGTTTGAAGCTAGAGAA--CCACGGGAACTGACTGTGGTCC	1013
Db	1403	AAATGCAAGTCTGTACGAGTTTGAAGCTAGAGAA--CCACGGGAACTGACTGTGGTCC	1462
QY	1014	AGGAGAGAGCTGGAGGTTCTGAGCACAGCAAGCGGTGGTGGTGAAGAAATGAGG	1073
Db	1463	AGGAGAGAGCTGGAGGTTCTGAGCACAGCAAGCGGTGGTGGTGAAGAAATGAGG	1522
QY	1074	CGGAGCGGAGCGGCTACATTTCCAGCAACATCTGGAGCCCTTACAGCCGGGACCCCTG	1133
Db	1523	CGGAGCGGAGCGGCTACATTTCCAGCAACATCTGGAGCCCTTACAGCCGGGACCCCTG	1582
QY	1134	GGACCCAGGCGCAGTCA-CCCTTCGGGTTCCAACTGCTTCCAGTTCAGTTCGAGGCGCTGAA	1192
Db	1583	GGACCCAGGCGCAGTCA-CCCTTCGGGTTCCAACTGCTTCCAGTTCAGTTCGAGGCGCTGAA	1642
QY	1193	GAGTCCACAGCTGGTGCAGGAGAGACTTCTCCACTCCACGCTGAGGACACTTGGG	1252
Db	1643	GAGTCCACAGCTGGTGCAGGAGAGACTTCTCCACTCCACGCTGAGGACACTTGGG	1702
QY	1253	TCCTGACGCGGG--AGCCAGCTACTTCGCA-TAAGACCTGGGGAGGTAC--AGATGCTAT	1307
Db	1703	TCCTGACGCGGGAGCCAGCTACTTCGCAATTAAGACCTGGGGAGCTACCGAGTCTAT	1762
QY	1308	GTCACAGGAGGCCCCAC---GAATCTGTCCCGGCTGGAGGCTGTGAGAGGATGC-TG	1363
Db	1763	GTCACAGGAGGCCCCAC---GAATCTGTCCCGGCTGGAGGCTGTGAGAGGATGC-TG	1822
QY	1364	GGATAGCCCTTAGGCACACGCTTAGACACCTCCCAAGAACCCGCGCTGATGCAAG	1423
Db	1823	GGATAGCCCTTAGGCACACGCTTAGACACCTCCCAAGAACCCGCGCTGATGCAAG	1882
QY	1424	ATGGCAGATCTGATCCCATTTAGAGCCCGGAGAAATTCCTCTTCTGGATCCCGTTTGCAG	1483
Db	1883	ATGGCAGATCTGATCCCATTTAGAGCCCGGAGAAATTCCTCTTCTGGATCCCGTTTGCAG	1942
QY	1484	CAAAACCCACAC--CCAGCTCACACAGCAAAACAAATGAGAGCCGCGAGGCTGAAGC	1541
Db	1943	CAAAACCCACACCTCCAGCGCTCACAGCAAAACAAATGAGAGCCGCGAGGCTGAAGC	2002
QY	1542	AAACAGTGTCCCTTCTGGCTGTCTTGGAGCTCCCGAGTCCCAAGAACTCTTATTTTACCT	1601
Db	2003	AAACAGTGTCCCTTCTGGCTGTCTTGGAGCTCCCGAGTCCCAAGAACTCTTATTTTACCT	2062
QY	1602	CTTTCCCAAACTGGAGCATTTATGCTAGGCTTGTGCAAGAACTCTGTCAGTCCCTCTCC	1661
Db	2063	CTTTCCCAAACTGGAGCATTTATGCTAGGCTTGTGCAAGAACTCTGTCAGTCCCTCTCC	2122
QY	1662	TTCTCAATAAAGCATCTTCAAGCTTGTAAAAAAA	1698
Db	2123	TTCTCAATAAAGCATCTTCAAGCTTGTAAAAAAA	2159
RESULT 6			
BC014734		2275 bp	mrna
LOCUS	BC014734		linear
			ROD 07-AUG-2002

DEFINITION	Mus musculus, Similar to hypothetical protein FLJ21522, clone MGC:25893 IMAGE:4218079, mRNA, complete cds.
ACCESSION	BC014734
VERSION	BC014734.1
KEYWORDS	GI:15928516
SOURCE	MGC.
ORGANISM	house mouse.
REFERENCE	Mus musculus. Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 2275)
TITLE	Strausberg, R.
JOURNAL	Direct Submission Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 31 Row: b Column: 17 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein. Location/Qualifiers 1..2275 /organism="Mus musculus" /db_xref="taxon:10090" /map="FVB/N" /clone="MGC:25893 IMAGE:4218079" /tissue_type="Colon, normal. 5 month old male mouse." /clone_lib="NCI_CGAP_Co24" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" 150..1952 /codon_start=1 /product="Similar to hypothetical protein FLJ21522" /protein_id="AAH14734.1" /db_xref="GI:15928517" /translation="MSRPSRSATYLHKEYSOSMASEPTLLQHRVHLMTCIKLGTORV REPDKALQKQEMDAGRWSDLFQVRDGLHLLDIETKEELDSYRLDNKAIKDA LNTCSYNSILTVQESGLPSTLLFCQEVGABQLRTSLQKALEELEERPFVGH HPSDRWKGPPLERPLPQQAPPLEQRFSPHFPPEQPHNMTSEKISPSSSSLTHY PSARENGFTLPPPRAPSPEDPERDEVLNHLRDLTELFAGKLKVEQARNSHKTK LGRKKKSKNGITQAEYIDCFQIKLSNLLGKLALRWQETSAPFVGLIFQTLKFL SQCPAGLPAKVISPELLTPKADLQSLSPEDLTKWSILGTSWTTSWADWTGSEPPP YQPTFDQWQIPQPSMPTITNQDSISLGRSMRSSLHFPDRPEYNNHPEYEDSNLPL SSPSGRAALKMVLVEFARNAGELTVAQGEILVLDQSKRWLVKKNKNEAGLTGIPS NLEPLPAGAPRGHPQSFRAPMLRLSSKPEVTAWLQAEFSTVTVRTLGLMSGLPS LHMPPGELQMLCPQEPAPRIQARLDLAVRRMLGTH"
BASE COUNT	620 a 651 c 565 g 439 t
ORIGIN	
Query Match	53.7%; Score 918; DB 10; Length 2275;
Best Local Similarity	74.4%; Pred. No. 1.4e-237;
Matches 1294; Conservative	0; Mismatches 396; Indels 49; Gaps 9;

/gene="Eps8R3"
/note="Eps8 Related protein 3"
/codon_start=1
/product="epidermal growth factor receptor pathway
substrate 8 related protein 3"
/protein_id="AAU76121.1"
/db_xref="GI:18655339"
/translation="MSRPSRAIYLHRKEYSQSMASEPFLQHRVHLMTCIKGTQVR
RPFKDALQIQMDAGRWVSODLFQVRDGLHLDIETKEELDSYRUDNKAIDVA
LNTCSVNSILSVTVQESGLPGISTLLFCQVEYGAELRTSLOKALEEERPRGVH
HPSODRWKGPPLERPLIQOAPLEOFSEPHREFPEQPHNMTSERSISPSRSLTHY
PARENGTFLPPPRAPSPEDPERDEVLNHVLRDIELFAGKLEKVEQARNHRTK
LGRKKKNGITQABYIDCFQIKLSFNLLKLRMOETSAPEFVLIQFLFKPIL
SQCEPALPKAIVISPLLPKALDILQSCLSPPEDTLWKSLSGTSWTTSWADWTGSEPP
YOPTFYDQWQIIPQPSMPITNQDSISLRGSRMSLSLHPRODEPNHNPEYEDSNLPL
SSPSGRKALMOVLVEFARNALQELTVAGQLELVLDQSKRWLVKNEAGLTGVIPI
NILEPLAPACAPRHOPSPFRAPMLRLSSKPEVTAWLQAFNEFTVTVRTLGSLMSQ
LHMRGELQMLCPQEAAPRIQARLDVARMGLMTH"

BASE COUNT 648 a 659 c 574 g 448 t
ORIGIN

Query Match 53.4%; Score 913.2; DB 10; Length 2329;
Best Local Similarity 74.2%; Pred. No. 2.8e-236;
Matches 1291; Conservative 0; Mismatches 399; Indels 49; Gaps 9;

Qy 1 GGCAGAGCGACTGAAGACCGCTGCGAGAAGGCTCTGGAGGAAGAGCTGGAGCAAGACC 60
Db 582 GGCAGAGCGACTGAGGACCAAGCTGCGAGAAGGCTCTGGAAGAAGAGCTAGAAGAAGACC 641

Qy 61 TCGACTTGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGCTATGAAAGGCC 120
Db 642 TCGATTGGAGTCCATCATCAAGCCAGGACAGATGGAAGGGGCGCTCTCTGAAAGGCC 701

Qy 121 GCTCCCTATGAGCAGCGACGCTATCTGAGCGGGGATC-----CC 162
Db 702 ACTCCCTATACAGCGCCACCCCTCTGAGCAGAGAGTCTCTCCAGAGCATAGGTTTC 761

Qy 163 TCCAGAACAGCCCAACAGAGCCCTTAGACACAGCTCCCAACCATCCCAAGGCCCT 222
Db 762 TCCAGAACAGCCACATAACATGACCTCAGAACGACGATCGCCCATCTCAAGTCTCT 821

Qy 223 GCCAGGCCACAGTGCCTCCGAGAACCAAGTGGCTTTACTCTG---CCTCCTCAAGGG 279
Db 822 GACACACTCCCAAGTCCCGGAGAACCAATGGCTTCACTCTGCTCTCTCTCCGAGGG 881

Qy 280 GTCTCTCTCCCGAGGACCCAGAGAGGAGGAGTGTGAACCATCTCTCAAGGGA 339
Db 882 TGTCTCTCTCCCGAGGACCCAGAGAGGATGAGAGGTGCTGAACCATCTCTCAAGGA 941

Qy 340 CATTGAGCTGTTTCATGGGAAGCTGGAGAAGGCCAGGCAAGAACAGCAGGAGGAAGAA 399
Db 942 CATTGAGCTGTTTCATGGGAAGCTGGAGAAGGTCAGGAGTCCAGGCAAGAACAGTCA 1001

Qy 400 A-----TTTGGGAAAAAACAGAGACAGGAGGCTCTCACCAGGCAAGTACATGA 453
Db 1002 AAAACTTTGGGAAAAAAGAAAGCTAAATAATGGGATAACCAAGCAGAGTACATGA 1061

Qy 454 CTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCGGAGGCTGCCACCTGGCTGAA 513
Db 1062 CTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCGGAGGCTGCCACCTGGGATGCA 1121

Qy 514 GGAGACAAGTGCCTGAGCTTACACATCTCTTCAAGTCCCTGAACCTTCACTCTGGC 573
Db 1122 GGAACAAGCGCCCTGAGTTCGTGGCCCTCATCTTCAAAACCTGAAATTTATCTATC 1181

Qy 574 CAGTGGCCCTGAGCTGGCTAGCAGCCCAAGTATCTACCCCTCTCTACCCCTAAAGC 633
Db 1182 TCAGTGGCCCTGAGCTGGCTTCCAGCAAGTATTTCCCTCTCTCTACCCCTAAAGC 1241

Qy 634 TATCAACCTCTACAGTCTCTTAAGCCCACTGAGAGTAACTTTGGATGGGTTGGG 693
Db 1242 AATAGACCTCTGAGTCTCTTAAGCCCACTGAGAGTAACTTTGGATGGGTTGGG 1301

RESULT 8
AX407643/c
LOCUS
DEFINITION

Sequence 290 from Patent W00229103.

401 bp DNA linear PAT 14-JUN-2002

Qy 694 CCAGCCTGGACACACTAGCCGGGCGGCGACTGGACAGCGCATGAGCCCTGCCCTACCAACC 753
Db 1302 CACATCTGGACACCACTGGCTGGCTGACTGGACAGCGAGTGAACCGCACCTTACCAACC 1361

Qy 754 CACATCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCAAGCACCCTTATGATACCA 813
Db 1362 CACATCTCAGATGACTGGCGAGATCCACAGCCCGCTCCATGATGCGCCATAACAACCA 1421

Qy 814 GGACCTGTTTCCCTTTCGGGGGGAAGTATAGTTAGGAGCAGCACTTCACACTTTCTCTCA 873
Db 1422 GGATTCGATTTCCCTCC-----GAGTTCTAGATGAGGAGCGCTACATTTCCCTCG 1475

Qy 874 GGAGAAGACACAAACATGACCTCAGCCTGGGGACCCCAACTCCAGGCCCTCCAGCCC 933
Db 1476 GGATGAGCCATACAACCAATAACTGAGGAGTCAAACTCCCACTCTCCAGCCC 1535

Qy 934 CAACCTGCCACCCAGCCCTGAAATGCAAGTCTGTACAGTTTGAACCTAGGACCC 993
Db 1536 CAGCCTGGCAGAGCAGCCCTGAAATGCAAGTCTGTACAGTTTGAACCAAGGAATGC 1595

Qy 994 ACGGGAACCTGACTGTGTCCAGGGAGAGAAGCTGGAGGTTCTGGACCACAGCGGGTG 1053
Db 1596 CCAGGAACCTGACTGTGCGACAGGGGAGATTTCTGGAGTTCTGATCAGAGCAAGCGATG 1655

Qy 1054 GTGGCTGGTGAAGATGAGGGGAGCGGCTGACTATTCACAGCAACATCTCTGGAGCC 1113
Db 1656 GTGGCTGGTGAAGAACGAGCAGGAGTGTGTTTACATTTCCCAAGCAATCTCTGGAGCC 1715

Qy 1114 CCTACAGCCGGGACCCCTGGGACCCAGGCCAGTCAACCTCTCGGTTTCCAATGCTTCG 1173
Db 1716 CTTGCCAGCTGGAGCCCGGAGGCCACCAACCATCTTTAGGGCCCAATGCTTCG 1775

Qy 1174 ACTTAGCTCAGGCGCTGAAGAGTTCAGAGTTCAGGCTGCGAGCGAGAACTTCTCCACTGC 1233
Db 1776 CTTGAGCTCAAGCGCTGAGGAGTTCAGGCTGCTGCTACAGCAGAGAACTTCTCCACTGT 1835

Qy 1234 CACGGTGAGGACACTTGGTCTCCTGAGCGGGAGCCAGCTACTTCGCAATAAGACCTGGGGA 1293
Db 1836 CACTGTAAGCACTCTCGGTCCTTAATGGGAAGCAGCTACTTACATGAGACTGGGGA 1895

Qy 1294 GCTACAGATCTATGCTCCAGGAGGCCCAAGTCTCTCCGGCTGGAGCTGTCTAG 1353
Db 1896 GCTGCAAGTCTGTGTCACAGGAGGCTCCAGGATCCAGGGCCCGGCTGGAGCTGTCTAG 1955

Qy 1354 AAGGATCTGGGATTAAGCCCTTAGGCACAGCTTAGACACCTCCAAAGAACACAGG-CCCC 1412
Db 1956 AAGGATCTGGGATGACTCATTTAGAGCAGCTCGATACCTCTAGAGCAAGGAGCTC 2015

Qy 1413 GCTGATGCAAGATGGCAGATCTGATACCCA--TTAGAGCCCCGAGAAATTCCTTCTGGAT 1471
Db 2016 TCATGTGCAAGATGGAGTATTTGTTACCCAGTTAGAGCCCTCGAAGTCTCTT----- 2069

Qy 1472 CCGAGTTTGAGCAAAACCCACACCCAGCTCAGACAGCAAAACAATGACAGCGCCAG 1531
Db 2070 -TCAGTTAGTGGCAAGCCACATATCTTGGATCAAAAGCAAAAGAGATGAACAGTCCAG 2128

Qy 1532 AGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCTTCCCAAGTAAACCACTATT 1591
Db 2129 AACTGAGGCAAAATGGTACCTCATTTGGTTG---GAGACCTTCAACCGTTA----CTATT 2181

Qy 1592 TATTTTACCTCTTTTCCCAAAACCTGGAGCAATTTATGCTTAGGCTTGTCAAGAATCTGTTC 1651
Db 2182 TATTTTACCCACACACCTTAAGTTGGCATACTTATTTCCAGTCTTACTCTGATCTCTTTA 2241

Qy 1652 GTCCCTCTCTCTCAATAAAGCAATCTCAAGCTTGTAAATAAATAAATAAANGATAAA 1710
Db 2242 GTCCCAATCTCGCAATAAAGCAATCGTTTAACTTGTATGAAATAAATAAATAAATAA 2300

repeat_region /note="LRR7 repeat: matches 1. .450 of consensus"
15622. .15992
/note="MER4-internal repeat: matches 1502. .1866 of
consensus"
15995. .16388
/note="MER4-internal repeat: matches 836. .1268 of
consensus"
16466. .16505
/note="MER4-internal repeat: matches 687. .726 of
consensus"
16510. .16805
/note="AluSc repeat: matches 1. .296 of consensus"
16815. .16886
/note="MER4-internal repeat: matches 751. .824 of
consensus"
17226. .17859
/note="MER4-internal repeat: matches 1. .692 of consensus"
17863. .18321
/note="MER4A2 repeat: matches 1. .504 of consensus"
18590. .18710
/note="L2 repeat: matches 2601. .2725 of consensus"
18829. .19149
/note="MLT1I repeat: matches 16. .363 of consensus"
19332. .19625
/note="AluX repeat: matches 1. .294 of consensus"
19659. .20116
/note="LRR2 repeat: matches 2. .449 of consensus"
20173. .20260
/note="MIR repeat: matches 2. .91 of consensus"
20328. .20469
/note="L2 repeat: matches 2131. .2282 of consensus"
21358. .21860
/note="MLT1E repeat: matches 1. .539 of consensus"
21876. .22075
/note="MER61D repeat: matches 342. .743 of consensus"
22851. .22889
/note="LIM3e repeat: matches 515. .556 of consensus"
23193. .23489
/note="AluJo repeat: matches 1. .296 of consensus"
23563. .23648
/note="LIME repeat: matches 5683. .5757 of consensus"
23649. .23955
/note="AluSc repeat: matches 1. .307 of consensus"
23956. .24093
/note="LIME repeat: matches 5497. .5683 of consensus"
24386. .24441
/note="MER41A repeat: matches 400. .452 of consensus"
26554. .26853
/note="Charlie4a repeat: matches 227. .508 of consensus"
26854. .27170
/note="AluJb repeat: matches 1. .312 of consensus"
27171. .27348
/note="Charlie4a repeat: matches 2. .227 of consensus"
27745. .28699
/note="CpG Island"
29257. .29603
/evidence=not_experimental
/note="L2 repeat: matches 2424. .2750 of consensus"
29892. .30198
/note="L2 repeat: matches 1322. .1655 of consensus"
30614. .30739
/note="MIR repeat: matches 137. .261 of consensus"
30741. .30788
/note="24 copies 2 mer tt 75% conserved"
30802. .36956
/note="LipA4 repeat: matches 6. .6146 of consensus"
36958. .37084
/note="MIR repeat: matches 23. .151 of consensus"
41990. .42041
/note="L2 repeat: matches 2652. .2703 of consensus"
42995. .43450
/note="LIMB4 repeat: matches 5731. .6181 of consensus"
43538. .43843
repeat_region

repeat_region /note="AluJb repeat: matches 1. .290 of consensus"
44248. .44545
/note="AluSq repeat: matches 7. .306 of consensus"
47160. .47197
/note="19 copies 2 mer gt 100% conserved"
50429. .50483
/note="L2 repeat: matches 2443. .2497 of consensus"
50653. .50853
/note="MIR repeat: matches 18. .233 of consensus"
53782. .53851
/note="L2 repeat: matches 2685. .2749 of consensus"
53853. .54465
/note="L2 repeat: matches 1749. .2413 of consensus"
54631. .55246
/note="LRR5 repeat: matches 1. .623 of consensus"
55326. .55629
/note="LRR5 repeat: matches 675. .969 of consensus"
55714. .55843
/note="L2 repeat: matches 2387. .2516 of consensus"
55897. .55938
/note="21 copies 2 mer tg 81% conserved"
56457. .56528
/note="L2 repeat: matches 2387. .2459 of consensus"
56728. .57841
/note="L2 repeat: matches 1389. .2609 of consensus"
59949. .60010
/note="31 copies 2 mer gt 80% conserved"
60563. .60721
/note="MER5B repeat: matches 18. .177 of consensus"
61673. .61733
/note="MIR repeat: matches 104. .164 of consensus"
61971. .61994
/note="MIR repeat: matches 117. .140 of consensus"
62581. .62627
/note="MIR repeat: matches 200. .240 of consensus"
62628. .62901
Query Match 19.0%; Score 324.4; DB 9; Length 133451;
Best Local Similarity 99.7%; Pred. No. 1.6e-76;
Matches 325; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1366 GATAAGCCCTTAGCACCAGCTTAGACACCTCCAAGAACCCAGGCCCGCTGATGCAAGAT 1425
Db 38418 GATAAGCCCTTAGCACCAGCTTAGACACCTCCAAGAACCCAGGCCCGCTGATGCAAGAT 38359
QY 1426 GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTTCGTGATCCCAGTTGCAGCA 1485
Db 38358 GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTTCGTGATCCCAGTTGCAGCA 38299
QY 1486 AACCCACACCCAGCTCACACAGCAAAACAAATGGAGCCCGCAGAGGCTGAAGCAAAC 1545
Db 38298 AACCCACACCCAGCTCACACAGCAAAACAAATGGAGCCCGCAGAGGCTGAAGCAAAC 38239
QY 1546 AGTGTCCTCTTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTTATTTTACCTCTTT 1605
Db 38238 AGTGTCCTCTTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTTATTTTACCTCTTT 38179
QY 1606 CCCAACCTGGAGCATTTATGCCTAGGCTTGTCAGAAATCTGTTCACTCCTCTCTCTCT 1665
Db 38178 CCCAACCTGGAGCATTTATGCCTAGGCTTGTCAGAAATCTGTTCACTCCTCTCTCTCT 38119
QY 1666 CAATAAAGCATCTTCAAGCTTTGTAA 1691
Db 38118 CAATAAAGCATCTTCAAGCTTTGTCA 38093
RESULT 10
AC092860 162480 bp DNA linear HTG 19-AUG-2002
LOCUS Pan troglodytes clone rpd43-125j14, WORKING DRAFT SEQUENCE, 5
DEFINITION ordered pieces.
AC092860
AC092860
VERSION AC092860.16 GI:22297381

KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE chimpanzee.
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
REFERENCE 1 (bases 1 to 162480)
AUTHORS Jiang.X., Song,L., Eichler,E. and Roe,B.A.
TITLE Pan troglodytes BAC Clone rp43-125j14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 162480)
AUTHORS Jiang.X., Song,L. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 162480)
AUTHORS Jiang.X., Song,L., Eichler,E. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (19-AUG-2002) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 19, 2002 this sequence version replaced gi:22218546.

Center: Department of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2375: contig of 2375 bp in length
* 2376 16668: contig of 14193 bp in length
* 2476 16669 16768: gap of unknown length
* 16669 16768: gap of unknown length
* 16769 35334: contig of 18766 bp in length
* 35335 35634: gap of unknown length
* 35635 92812: contig of 57178 bp in length
* 92813 92913: gap of unknown length
* 92913 162480: contig of 69568 bp in length.
FEATURES
source
1..162480
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone_lib="rp43-125j14"
/clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
BASE COUNT 42683 a 39383 c 38740 g 41229 t 445 others
ORIGIN
Query Match 18.9%; Score 322.8; DB 2; Length 162480;
Best Local Similarity 99.4%; Pred. No. 4.5e-76;
Matches 324; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1366 GATAGCCCTTAGGCACCGCTTAGACACCTCCCAAGAACACGAGCCCGCTGATGCAAGAT 1425
DB 44411 GATAGCCCTTAGGCACCGCTTAGACACCTCCCAAGAACACGAGCCCGCTGATGCAAGAT 44470
QY 1426 GGCAGATCTGATACCCATTAGAGCCCGCCAGAAATTCCTCTTGGATCCAGCTTGCAGCA 1485
DB 44471 GGCAGATCTGATACCCATTAGAGCCCGCCAGAAATTCCTCTTGGATCCAGCTTGCAGCA 44530
QY 1486 AACCCACACCCCGAGCTCACACAGCAAAACAAATGGACAGGCCCGCCAGAGCGCTCAAGCAAAAC 1545
DB 44531 AACCCACACCCCGAGCTCACACAGCAAAACAAATGGACAGGCCCGCCAGAGCGCTCAAGCAAAAC 44590
QY 1546 AGTGTCCCTCTTGGCTGTGTGGAGCCCTCCCGAGTAACACCTATTATTATTACCTCTTT 1605
DB 44591 AGTGTCCCTCTTGGCTGTGTGGAGCCCTCCCGAGTAACACCTATTATTATTACCTCTTT 44650

QY 1606 CCCAAACCTGGAGCATTTATGCTTAGCTTGTCAAGAAATCTGTCAGTCCCTCTCTCTTCT 1665
DB 44651 CCCAAACCTGGAGCATTTATGCTTAGCTTGTCAAGAAATCTGTCAGTCCCTCTCTCTTCT 44710
QY 1666 CAATAAAAGCATCTTCAAGCTTGTAA 1691
DB 44711 CAATAAAAGCATCTTCAAGCTTGTCA 44736
RESULT 11
G25672
LOCUS human STS EST47775, sequence tagged site. 224 bp DNA linear STS 02-JUN-1996
ACCESSION G25672
VERSION G25672.1 GI:1347904
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens STSS derived from sequences in dbEST and the Unigene collection.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 224)
AUTHORS Hudson,T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically Mapped STSS
JOURNAL Unpublished (1995)
COMMENT Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: GTTTCAGCAAAACCCAC
Primer B: TCCAGGTTGGGAAAGAGGT
STS size: 142
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3
Derived from dbEST (genbank accession D25742).
FEATURES
source
Location/Qualifiers
1..224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="406.4 cR from top of Chr1 linkage group"
STS
primer_bind 8..25
BASE COUNT 61 a 67 c 37 g 55 t 4 others
ORIGIN
Query Match 12.8%; Score 218.4; DB 11; Length 224;
Best Local Similarity 97.8%; Pred. No. 5.5e-48;

Matches 219; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1469 GATCCAGTTTGAGCAAAACCCACACCCAGCTCACAGCAAAACAAATGGACAGGC 1528
|||||
Db 1 GATCCAGTTTGAGCAAAACCCACACCCAGCTCACAGCAAAACAAATGGACAGGC 60

QY 1529 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACACCT 1588
|||||
Db 61 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACACCT 120

QY 1589 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 1648
|||||
Db 121 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 180

QY 1649 TCAGTCCCTCTCCTCTCAATAAAGCATCTTCAAGCTTGTAAA 1692
|||||
Db 181 TCAGTCCCTCTCCTCTCAATAAAGCATCTTCAAGCTTGTAAA 224

RESULT 12
G41866
LOCUS 224 bp DNA linear STS 29-SEP-1998
DEFINITION SHGC-32779 Human Homo sapiens STS genomic, sequence tagged site.
ACCESSION G41866
VERSION G41866.1 GI:3668199
KEYWORDS STS.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 224)
Myers,R.M.
Human STSs (1998)
Unpublished (1998)
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@hgc.stanford.edu
Primer A: GTTTCGACAAACCCAC
Primer B: TCCAGGTTTGGAAAGAGT
STS size: 142
PCR Profile:
Initial Incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifit Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
Prepared with primer pairs provided by Sandoz, derived from D25742
-- Washington University/Merck EST sequence.
Location/Qualifiers
1..224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1"

STSS
8..149
primer_bind 8..25
primer_bind complement(130..149)
BASE COUNT 61 a 67 c 37 g 55 t 4 others
ORIGIN

Query Match 12.8%; Score 218.4; DB 11; Length 224;
Best Local Similarity 97.8%; Pred No. 5.5e-48;
Matches 219; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1469 GATCCAGTTTGAGCAAAACCCACACCCAGCTCACAGCAAAACAAATGGACAGGC 1528
|||||
Db 1 GATCCAGTTTGAGCAAAACCCACACCCAGCTCACAGCAAAACAAATGGACAGGC 60

QY 1529 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACACCT 1588
|||||
Db 61 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACACCT 120

QY 1589 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 1648
|||||
Db 121 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 180

QY 1649 TCAGTCCCTCTCCTCTCAATAAAGCATCTTCAAGCTTGTAAA 1692
|||||
Db 181 TCAGTCCCTCTCCTCTCAATAAAGCATCTTCAAGCTTGTAAA 224

RESULT 13
AR044148
LOCUS 197 bp DNA linear PAT 29-SEP-1999
DEFINITION Sequence 40 from patent US 5817479.
ACCESSION AR044148
VERSION AR044148.1 GI:5965613
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 197)
AUTHORS Au-Young,J., Bandman,O., Hawkins,P.R. and Wilde,C.G.
TITLE Human Kinase homologs
JOURNAL Patent: US 5817479-A 40 06-OCT-1998;
FEATURES Location/Qualifiers
source 1..197
/organism="unknown"
BASE COUNT 50 a 52 c 63 g 32 t
ORIGIN

Query Match 9.1%; Score 155.6; DB 6; Length 197;
Best Local Similarity 96.0%; Pred. No. 5.8e-31;
Matches 192; Conservative 0; Mismatches 4; Indels 4; Gaps 3;

QY 932 CCAAAACCTGCCAGCCAGCCCTGAAATGCAAGTTTGTACAGTTTGAAGCTAGGAAC 991
|||||
Db 1 CCAAAACCTGCCAGCCAGCCCTGAAATGCAAGTTTGTACAGTTTGAAGCTAGGAAC 60

QY 992 CCACGGGAAGTACTGTGTCAGGAGAGAGCTGGAGTCTTGCACACACACAGCCG 1051
|||||
Db 61 CCACGGGAAGTACTGTGTCAGGAGAGAGCTGGAGTCTTGCACACACACAGCCG 119

QY 1052 TGGTGGCTGGTGAAGAAATGAGCGGAGCGGCTACATTTCAAGCAACATCTGGAG 1111
|||||
Db 120 TGGTGGCTGGTGAAGAAAT-AGCGGGAGCGGAGCGGCTACATTTCAAGCAACATCT--TGGG 176

QY 1112 CCCTTACAGCCGGGACCC 1131
|||||
Db 177 CCCTTACAGCCGGGACCC 196

RESULT 14
AC018461/c
LOCUS 113285 bp DNA linear ROD 15-JUN-2002
DEFINITION Mus musculus chromosome 3 clone mgs1-192g21 strain 129/SvJ,

```
complete sequence.
AC018461
VERSION AC018461.30 GI:21427705
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 113285)
JOURNAL Ray, L., Qi, S., Loh, P., Perng, T., Carson, M., Pearson, W. and Roe, B.A.
REFERENCE Mus musculus BAC Clone mgs1-192q21
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 113285)
TITLE Ray, L., Qi, S., Loh, P., Pearson, B. and Roe, B.A.
JOURNAL Direct Submission
AUTHORS Submitted (11-DEC-1999) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 113285)
AUTHORS Ray, L., Qi, S., Loh, P., Pern, T., Carson, M., Pearson, W. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 4 (bases 1 to 113285)
AUTHORS Ray, L., Qi, S., Loh, P., Pern, T., Carson, M., Pearson, W. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Jun 15, 2002 this sequence version replaced gi:21328313.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES             Location/Qualifiers
     source          1. 113285
                     /organism="Mus musculus"
                     /strain="129/SvJ"
                     /db_xref="taxon:10090"
                     /chromosome="3"
                     /clone="mgs1-192q21"
                     /clone_lib="Genome Systems mouse ES cell BAC libraries I
                     and II"
BASE COUNT      28787 a 26952 c 27495 g 30051 t
ORIGIN
Query Match          6.9%; Score 117.6; DB 10; Length 113285;
Best Local Similarity 75.0%; Pred. No. 1.9e-20;
Matches 147; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 837 GAAGTCATAGTTAGGGAGCAGCTCACACTTCTCTCAGAGAGACACACACCATGACC 896
DB 43033 GAGGTTCTAGTAGGAGAGCAGCTACATTTCCCTCGGATGAGCCATACACCATACC 42974
QY 897 CTCACCTCGGGAGCCCAACTCCAGCCCTCCAGCCCCCAAACTCCCGAGCCGCTGA 956
DB 42973 CTGAATACGAGGACATCAACCTCCCACTCCAGCCCCCAGCCCTGGCAGAGCAGCCCTGA 42914
QY 957 AAATGCAAGTCTTTACGAGTTTGAAGTAGGAACCCAGCGGAAGTACTGTGGTCCAGG 1016
DB 42913 AAATGCAAGTCTGTACGAGTTTGAAGCAAGGAATGCCAGGAAGTACTGTGGCAGAGG 42854
QY 1017 GAGAGAAGCTGGAGGT 1032
DB 42853 GGGAGATTCTGGAGGT 42838
RESULT 15
AC079042
LOCUS AC079042 198459 bp DNA linear HTG 20-AUG-2002
DEFINITION Mus musculus chromosome 3 clone rp23-313f7 strain C57BL/6J, WORKING
-----
DRAFT SEQUENCE, 2 ordered pieces..
AC079042
VERSION AC079042.27 GI:22325017
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 198459)
JOURNAL Sloan, D., Hua, A., Hong, X., Pearson, W. and Roe, B.A.
REFERENCE Mus musculus BAC Clone rp23-313f7
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 198459)
TITLE Sloan, D., Hua, A., Hong, X., Pearson, W. and Roe, B.A.
JOURNAL Direct Submission
AUTHORS Submitted (17-AUG-2000) Department Of Chemistry And Biochemistry,
JOURNAL The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 3 (bases 1 to 198459)
AUTHORS Sloan, D., Hua, A., Hong, X., Pearson, W. and Roe, B.A.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On Aug 20, 2002 this sequence version replaced gi:22218516.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 57584: contig of 57584 bp in length
* 57585 57684: gap of unknown length
* 57685 198459: contig of 140775 bp in length.
FEATURES             Location/Qualifiers
     source          1. 198459
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /chromosome="3"
                     /clone="rp23-313f7"
                     /clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT      55177 a 47384 c 46269 g 49494 t 135 others
ORIGIN
Query Match          6.9%; Score 117.6; DB 2; Length 198459;
Best Local Similarity 75.0%; Pred. No. 2e-20;
Matches 147; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 837 GAAGTCATAGTTAGGGAGCAGCTCACACTTCTCTCAGAGAGACACACACCATGACC 896
DB 128837 GAGGTTCTAGTAGGAGAGCAGCTACATTTCCCTCGGATGAGCCATACCATACC 128896
QY 897 CTCAGCCTCGGGAGCCCCAACTCCAGCCCTCCAGCCCCCAAACTGCCAGCCGCTGA 956
DB 128897 CTGAATACGAGGACTCAAACTCCCACTCTCCAGCCCCCAGCCCTGGCAGAGCAGCCCTGA 128956
QY 957 AAATGCAAGTCTTTACGAGTTTGAAGTAGGAACCCAGCGGAAGTACTGTGGTCCAGG 1016
DB 128957 AAATGCAAGTCTTTACGAGTTTGAAGCAAGGAATGCCAGGAAGTACTGTGGCAGAGG 129016
QY 1017 GAGAGAAGCTGGAGGT 1032
DB 129017 GGGAGATTCTGGAGGT 129032
```

Search completed: February 25, 2003, 08:44:52
Job time : 3704 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 05:25:18 ; Search time 261 Seconds
(without alignments)
14754.472 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 1710

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1709	99.9	1710	21	AAZ95379 Human colon specif
2	1696.6	99.2	2220	22	AAF32641 Human cDNA encodin
3	1334.6	78.0	2159	24	ABA91311 Human colon specif
4	832.8	48.7	893	23	ABK43674 DNA encoding novel
5	830.8	48.6	872	22	AAH33515 Human colon cancer
6	829.6	48.5	872	22	AAH33515 cDNA encoding nove
7	829.6	48.5	872	23	AAK43957 DNA encoding novel
8	673.8	39.4	1265	22	AAZ95379 cDNA encoding nove
9	671.8	39.3	1282	23	ABK43673 DNA encoding novel

c	10	389	22.7	401	24	ABN93792	Gene #290 used to
	11	388.8	22.7	1388	22	AAK88048	Human digestive sy
	12	335.4	19.6	572	22	AAZ95379	cDNA encoding nove
	13	335.4	19.6	572	23	ABK43956	DNA encoding novel
	14	331	19.4	786	20	AAZ16535	Human gene expres
c	15	324	18.9	2643	22	AAZ16535	DNA encoding novel
	16	297.8	17.4	338	22	AAH34654	Human colon cancer
	17	218.4	12.8	224	16	AAZ22498	Human gene signatu
	18	180.2	10.5	481	22	AAH36539	Human colon cancer
	19	157.4	9.2	6784	22	AAK89660	Human digestive sy
	20	155.6	9.1	197	19	AAV33556	Clone 237002 CDNA
	21	119.2	7.0	216	21	AAK98311	Human colon cancer
	22	95.8	5.6	2457	22	AAH99536	Human protein enco
	23	93.2	5.5	1635	22	AAH24282	Human Ovi107 ovar
	24	86.4	5.1	2535	22	AAK94459	Human full-length
	25	68.8	4.0	595	21	AAK98568	Human colon cancer
	26	62.6	3.7	2922	23	ABL18887	Drosophila melanog
	27	62.6	3.7	3031	23	ABL07407	Drosophila melanog
	28	62.6	3.7	6013	23	ABL18686	Drosophila melanog
	29	62.6	3.7	8511	23	ABL07406	Human spliced tran
	30	60	3.5	60	24	ABN41901	DNA encoding novel
	31	56.4	3.3	1638	23	ABK43695	cDNA encoding nove
	32	56.4	3.3	1639	22	AAZ27064	Human ovarian and
	33	56.4	3.3	1661	22	ABA07672	Human ovary and
	34	56.4	3.3	1661	22	AAZ02553	Human reproductive
	35	56.4	3.3	1661	22	AAZ27477	cDNA encoding nove
	36	56.4	3.3	1661	23	ABK43975	DNA encoding novel
	37	56.4	3.3	1663	24	ABQ54695	Human ovarian anti
	38	56.4	3.3	1947	22	AAH14778	Human cDNA sequenc
	39	55	3.2	514	24	AAZ61502	Lung small cell ca
	40	53.8	3.1	1655	23	AAZ73902	DNA encoding novel
	41	49.8	2.9	3245	14	AAQ40730	Mouse eps8. Mus m
	42	49.8	2.9	10732	21	AAK10594	Gene encoding a su
	43	47.4	2.8	1457	22	AAZ92642	Human Nck-2 cDNA.
	44	47.4	2.8	2267	20	AAK84665	Human KDR signal t
c	45	46.6	2.7	9232	23	ABL20448	Drosophila melanog

ALIGNMENTS

RESULT 1
AAZ95379

ID AAZ95379 standard; DNA; 1710 BP.

AC AAZ95379;

XX

DT 01-JUN-2000 (first entry)

DE Human colon specific gene (CSG) SEQ ID NO:1.

DE Human colon specific gene (CSG) SEQ ID NO:1.

KW Human; colon specific gene; CSG; diagnosis; colon cancer; detection;
KW cytosolic; ds.
XX

OS Homo sapiens.

PN WO200007632-A1.

XX

PD 17-FEB-2000.

PF

XX 20-JUL-1999; 99WO-US16357.

XX

PR 04-AUG-1998; 98US-0095231.

XX

PA (DIAD-) DIADEXUS LLC.

XX

PI Sun Y, Recipon H, Macina RA;

XX

DR WPI; 2000-205579/18.

XX

PT Novel methods for diagnosing, monitoring, staging, imaging and treating
PT colon cancer by measuring the level of colon specific gene markers -
XX

PS Claim 6; Page 37-38; 42pp; English.

XX The present invention describes a method for diagnosing the presence of
CC colon cancer in a patient. The method comprises measuring levels of
CC colon specific gene markers (CSG) in cells, tissues or bodily fluids,
CC and comparing the measured levels of CSG with levels of CSG from a
CC normal human control, where an increase in measured CSG levels in the
CC patient versus control is associated with the presence of colon cancer.
CC AA295379 to AA295381 represent human CSG sequences. The method is used
CC to detect, monitor, stage or give a prognosis for colon cancer.
CC Antibodies against CSGs are used for detection or image localisation of
CC the CSGs. The antibodies can be conjugated to cytotoxic agent or drug and
CC used to treat colon cancer. The method of the invention is more accurate
CC than prior art clinical methods for staging colon cancer, because it
CC measures colon specific markers, and, unlike pathological staging
CC methods, do not depend on an invasive procedure.

XX Sequence 1710 BP; 437 A; 529 C; 444 G; 299 T; 1 other;

Query Match		99.9%;	Score 1709;	DB 21;	Length 1710;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1710;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	GCACAGCGACTGAAGACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACC	60		
Db	1	GCACAGCGACTGAAGACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACC	60		
QY	61	TCGACTTGGAGGCCCTTCACGACAGGCCAGGACAGATGGAGGGGCGCTGTATGGAAGGCC	120		
Db	61	TCGACTTGGAGGCCCTTCACGACAGGCCAGGACAGATGGAGGGGCGCTGTATGGAAGGCC	120		
QY	121	GCTCCTATGGAGCGCAGCGTATCTGGAGCCGGGGATCCCTCCAGAACAGCCCCACCA	180		
Db	121	GCTCCTATGGAGCGCAGCGTATCTGGAGCCGGGGATCCCTCCAGAACAGCCCCACCA	180		
QY	181	GAGGACCTTAGACAGAGCCTCCACCATCCCAAGGCCCTGCGCACACACCAAGTGC	240		
Db	181	GAGGACCTTAGACAGAGCCTCCACCATCCCAAGGCCCTGCGCACACACCAAGTGC	240		
QY	241	CCGAGAACCAAGTGCTTACTGCTCCTCCTCAAGGGGCTCCTCTCCCGGAGGACC	300		
Db	241	CCGAGAACCAAGTGCTTACTGCTCCTCCTCAAGGGGCTCCTCTCCCGGAGGACC	300		
QY	301	AGAGAGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGGGAA	360		
Db	301	AGAGAGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGGGAA	360		
QY	361	GCTGGAGAGGCCAGGCAAGAACACAGGAGGAAGAAATTTGGGAAAAAACAAGGA	420		
Db	361	GCTGGAGAGGCCAGGCAAGAACACAGGAGGAAGAAATTTGGGAAAAAACAAGGA	420		
QY	421	CCAGGAGGTCTCACCCAGGCACAGTACATTTGACTGCTTCCAGAGATCAAGTACAGCTT	480		
Db	421	CCAGGAGGTCTCACCCAGGCACAGTACATTTGACTGCTTCCAGAGATCAAGTACAGCTT	480		
QY	481	CAACCTCTCGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCTGAGCTCGTACA	540		
Db	481	CAACCTCTCGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCTGAGCTCGTACA	540		
QY	541	CATCTCTTCAAGTCCGTGAAGTTCATCTGCGCAGGTGCCCTGAGGCTGGCCTAGCAGC	600		
Db	541	CATCTCTTCAAGTCCGTGAAGTTCATCTGCGCAGGTGCCCTGAGGCTGGCCTAGCAGC	600		
QY	601	CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTAAG	660		
Db	601	CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTAAG	660		
QY	661	CCCACCTGAGAGTAACTTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCGGGCCGA	720		
Db	661	CCCACCTGAGAGTAACTTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCGGGCCGA	720		
QY	721	CTGGACAGCGGATGAGCCCTCTGCCCTACCAACCCACATCTTCAGATGACTGGCAACTTCC	780		
XX					

Db	721	CTGGACAGCGGATGAGCCCTCTGCCCTACCAACCCACATCTCTCAGATGACTGGCAACTTCC	780
QY	781	AGAGCCCTCCAGCCCAAGCACCTTAGGATACAGGACCCCTGTTTCCCTTCCGCGGGGAAG	840
Db	781	AGAGCCCTCCAGCCCAAGCACCTTAGGATACAGGACCCCTGTTTCCCTTCCGCGGGGAAG	840
QY	841	TCATAGGTTAGGAGCACCTCACACTTCTCAGGAGAAGACACAAACCATGACCCCTCA	900
Db	841	TCATAGGTTAGGAGCACCTCACACTTCTCAGGAGAAGACACAAACCATGACCCCTCA	900
QY	901	GCCTGGGACCCCACTCCAGCCCTCAGCCCAAACTGCCAGCCAGCCCTCAAAAT	960
Db	901	GCCTGGGACCCCACTCCAGCCCTCAGCCCAAACTGCCAGCCAGCCCTCAAAAT	960
QY	961	GCAAGCTTTGTACAGTTTGAAGCTTAGGAACCCACGGGAACCTGACTGTGGTCCAGGGAGA	1020
Db	961	GCAAGCTTTGTACAGTTTGAAGCTTAGGAACCCACGGGAACCTGACTGTGGTCCAGGGAGA	1020
QY	1021	GAACTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGAAGAATGAGCGGGAGC	1080
Db	1021	GAACTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGAAGAATGAGCGGGAGC	1080
QY	1081	GAGCGGCTACATTTCCAAGCAACATCTCGAGCCCTACAGCGGGGACCCCTGGGACCCA	1140
Db	1081	GAGCGGCTACATTTCCAAGCAACATCTCGAGCCCTACAGCGGGGACCCCTGGGACCCA	1140
QY	1141	GGCCAGTCACTCTCGGGTTCCAATGCTTCGACTTACGCTCGAGGCTTGAAGAGGTAC	1200
Db	1141	GGCCAGTCACTCTCGGGTTCCAATGCTTCGACTTACGCTCGAGGCTTGAAGAGGTAC	1200
QY	1201	AGACTGGCTGACGAGGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGGTCCCTGAC	1260
Db	1201	AGACTGGCTGACGAGGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGGTCCCTGAC	1260
QY	1261	GGGAGCCAGCTACTTCCGATTAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGC	1320
Db	1261	GGGAGCCAGCTACTTCCGATTAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGC	1320
QY	1321	CCACGAATCTGTCCTCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGGC	1380
Db	1321	CCACGAATCTGTCCTCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGGC	1380
QY	1381	ACCAGCTTAGACACCTCCAAGAACCCAGGCCCGCTGTATGCAAGATGGCAGATCTGATACC	1440
Db	1381	ACCAGCTTAGACACCTCCAAGAACCCAGGCCCGCTGTATGCAAGATGGCAGATCTGATACC	1440
QY	1441	CATTAGAGCCCGGAGAAATCTCTTCTGATCCCAAGTTTGCAGCAAAACCCACACCCAG	1500
Db	1441	CATTAGAGCCCGGAGAAATCTCTTCTGATCCCAAGTTTGCAGCAAAACCCACACCCAG	1500
QY	1501	CTCACACAGCAAAACAATGGACAGCCAGAGGCTGAAGCAAAACAGTGTCTCTTCTGGC	1560
Db	1501	CTCACACAGCAAAACAATGGACAGCCAGAGGCTGAAGCAAAACAGTGTCTCTTCTGGC	1560
QY	1561	TGTGTGGAGCCTCCCAAGTACCACTATTTATTTTACCTCTTTCCAAACCTGGAGCA	1620
Db	1561	TGTGTGGAGCCTCCCAAGTACCACTATTTATTTTACCTCTTTCCCAACCTGGAGCA	1620
QY	1621	TTTATGCTTACGCTTGTCAAGAACTCTGTTCAGTCCCTCTCTCTCAATTAAGCATCTT	1680
Db	1621	TTTATGCTTACGCTTGTCAAGAACTCTGTTCAGTCCCTCTCTCTCAATTAAGCATCTT	1680
QY	1681	CAAGCTTCTAAAAAATAAANGATATAA 1710	
Db	1681	CAAGCTTCTAAAAAATAAANGATATAA 1710	

RESULT 2
AAF32641
ID AAF32641 standard; cDNA; 2220 BP.
XX
AC AAF32641;
XX

22-MAR-2001 (first entry)
Human cDNA encoding intracellular signalling molecule INTRA4.
Human; intracellular signalling molecule; INTRA; immunosuppressive;
cytostatic; neuroprotective; nontropic; antiarteriosclerotic; cancer;
antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal;
antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS;
cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;
inflammatory disorder; Addison's disease; gastrointestinal disorder;
neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;
mental disorder; schizophrenia; anxiety; ss.
Homo sapiens.
WO200077040-A2.
21-DEC-2000.
16-JUN-2000; 2000WO-US16636.
16-JUN-1999; 99US-0139566.
17-AUG-1999; 99US-0149640.
09-NOV-1999; 99US-0164417.
(INCY-) INCYTE GENOMICS INC.
Yue H. Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;
Azimzai Y, Yang J, Reddy R, Lu DM;
WPT; 2001-025334/03.
P-PSDB; AAB64372.
New human intracellular signaling molecules, useful for the diagnosis,
prevention and treatment of cell proliferative, autoimmune,
inflammatory, neurological, gastrointestinal, reproductive and
developmental disorders -
Claim 5; Page 162; 192pp; English.
Sequences AAF32638 - AAF32689 represent cDNA encoding human
intracellular signalling molecules INTRA1 - INTRA52, represented in
AAB64369 - AAB64420. Modulators of the intracellular signalling molecules
of the invention exhibit immunosuppressive; cytostatic; neuroprotective;
neuroleptic; antiarteriosclerotic; antiinflammatory; anti-HIV;
antihelminthic; and antiparkinsonian activity. INTRA polypeptides their
agonists and antagonists are useful for the treatment of a condition
associated with decreased or increased expression of functional INTRA.
Disorders associated with abnormal INTRA expression or activity include
cell proliferative disorders e.g. arteriosclerosis and cancers;
autoimmune or inflammatory disorders e.g. Addison's disease and acquired
immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,
protozoal and helminthic infections; gastrointestinal disorders e.g.
dysphagia and irritable bowel syndrome; neurological disorders e.g.
epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob
disease and mental disorders e.g. anxiety, schizophrenia and Tourette's
disorder. Antibodies immuno specific for the INTRA proteins may also be
useful in the diagnosis of the above disorders.
Sequence 2220 BP; 549 A; 696 C; 593 G; 382 T; 0 other;
Query Match 99.2%; Score 1696.6; DB 22; Length 2220;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1699; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 GGCAGAGCGACTGAAGACCGCTGCAGAGGCTCTCGAGGAGAGAGCTGGAGCAAGAGACC 60
Db 518 GGCAGAGCGACTGAAGACCGCTGCAGAGGCTCTCGAGGAGAGAGCTGGAGCAAGAGACC 577
Qy 61 TCGACTTGGAGGCTTCAGCCAGGCGCAGACAGATGATGAGGGGGCTGCTATGGAAGGCC 120
Db 578 TCGACTTGGAGGCTTCAGCCAGGCGCAGACAGATGATGAGGGGGCTGCTATGGAAGGCC 637

Qy 121 GCTCCCTATGGAGCAGCAGCGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCACCA 180
Db 638 GCTCCCTATGGAGCAGCAGCGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCACCA 697
Qy 181 GAGACCCCTAGAGCAGCAGCGCTCCACCATCCCAAGGCCCTGCCAGCCACACAGTGC 240
Db 698 GAGACCCCTAGAGCAGCAGCGCTCCACCATCCCAAGGCCCTGCCAGCCACACAGTGC 757
Qy 241 CCGAGAACCAAGTGCCTTTACTCTGCTCCCTCCCAAGGGGCTCTTCCCCCGAGGAGACC 300
Db 758 CCGAGAACCAAGTGCCTTTACTCTGCTCCCTCCCAAGGGGCTCTTCCCCCGAGGAGACC 817
Qy 301 AGAGAGGAGCAGGAGAGTGTCTGAACCATGTCTTAAGGAGACATTTGAGTGTTCATGGGAAA 360
Db 818 AGAGAGGAGCAGGAGAGTGTCTGAACCATGTCTTAAGGAGACATTTGAGTGTTCATGGGAAA 877
Qy 361 GCTGGAGAGGCCCCAGGCAAGACCCAGAGAGAGAGAAATTTGGGAAAAAACAAGGA 420
Db 878 GCTGGAGAGGCCCCAGGCAAGACCCAGAGAGAGAGAAATTTGGGAAAAAACAAGGA 937
Qy 421 CCAGGGAGGTCTCACCCAGGCGACAGTACATTTGAGTGTCTCCAGAGAGATCAAGTACAGCTT 480
Db 938 CCAGGGAGGTCTCACCCAGGCGACAGTACATTTGAGTGTCTCCAGAGAGATCAAGTACAGCTT 997
Qy 481 CAACCTCTGGGAGAGCTGGCCACTGGCTGAAGGAGACAAAGTCCCTTGAGCTCGTACA 540
Db 998 CAACCTCTGGGAGAGCTGGCCACTGGCTGAAGGAGACAAAGTCCCTTGAGCTCGTACA 1057
Qy 541 CATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGCAGC 600
Db 1058 CATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGCAGC 1117
Qy 601 CCAAGTGTCTCACCCCTCTCTCAACCTTAAGCTATCAACCTGTACAGTCTGTCTTAAG 660
Db 1118 CCAAGTGTCTCACCCCTCTCTCAACCTTAAGCTATCAACCTGTACAGTCTGTCTTAAG 1177
Qy 661 CCACCTTGAGAGTAACCTTTGGATGGGGTTGGGCCAGCCCTGGACCACTAGCCGGGCCGA 720
Db 1178 CCACCTTGAGAGTAACCTTTGGATGGGGTTGGGCCAGCCCTGGACCACTAGCCGGGCCGA 1237
Qy 721 CTGACAGGCGATGAGGCCCTTCCCTTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 780
Db 1238 CTGACAGGCGATGAGGCCCTTCCCTTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 1297
Qy 781 AGACCCCTCCAGCAACGACCCCTTAGATACAGGAGCCCTGTCTTCCCTTGGCGGGGAAG 840
Db 1298 AGACCCCTCCAGCAACGACCCCTTAGATACAGGAGCCCTGTCTTCCCTTGGCGGGGAAG 1357
Qy 841 TCATAGCTTAGGAGCAGCCTCACACTTTCCTCAGGAGAGACACACACCAACCATGACCTCA 900
Db 1358 TCATAGCTTAGGAGCAGCCTCACACTTTCCTCAGGAGAGACACACACCAACCATGACCTCA 1417
Qy 901 GCCTGGGGACCCCAACTTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGGAAT 960
Db 1418 GCCTGGGGACCCCAACTTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGGAAT 1477
Qy 961 GCAGTCTTGTACAGTTTGAAGCTAGAACCCAGGAACTGACTGTGGTCCAGGAGCA 1020
Db 1478 GCAGTCTTGTACAGTTTGAAGCTAGAACCCAGGAACTGACTGTGGTCCAGGAGCA 1537
Qy 1021 GAAGCTGGAGGTTCTGGACCCACAGAGCGGTGGTGGCTGGTGAAGATGAGCGGGGACG 1080
Db 1538 GAAGCTGGAGGTTCTGGACCCACAGAGCGGTGGTGGCTGGTGAAGATGAGCGGGGACG 1597
Qy 1081 GAGCGGTACATTCACAGCAACATCTCTGGAGCCCTTACAGCCGGGGACCCCTGGGACCCA 1140
Db 1598 GAGCGGTACATTCACAGCAACATCTCTGGAGCCCTTACAGCCGGGGACCCCTGGGACCCA 1657
Qy 1141 GGCCAGCTCACCTCTCGGGTTCCAAATGCTTGCATTTAGCTCGAGGCGCTGAGAGGTCAC 1200
Db 1658 GGCCAGCTCACCTCTCGGGTTCCAAATGCTTGCATTTAGCTCGAGGCGCTGAGAGGTCAC 1717

QY 1201 AGACTGGCTGCAGGAGAACTTCTCCACTCCCGGTGAGGACACTTGGTCCCTGAC 1260
DB 1718 AGACTGGCTGCAGGAGAACTTCTCCACTCCCGGTGAGGACACTTGGTCCCTGAC 1777
QY 1261 GGGAGCCAGCTACTTCCGATAGACCTGGGGAGCTACAGATGCTATGTCCACAGAGGC 1320
DB 1778 GGGAGCCAGCTACTTCCGATAGACCTGGGGAGCTACAGATGCTATGTCCACAGAGGC 1837
QY 1321 CCCAGCAATCTCTCCCGCTCGAGGCTCTCAGAGGATGCTGGGATAGCCCTTAGGC 1380
DB 1838 CCCAGCAATCTCTCCCGCTCGAGGCTCTCAGAGGATGCTGGGATAGCCCTTAGGC 1897
QY 1381 ACCAGCTTAGACACCTTCCAAAGAACAGGCCGCTGATGCAAGATGGCAGATCTGATACC 1440
DB 1898 ACCAGCTTAGACACCTTCCAAAGAACAGGCCGCTGATGCAAGATGGCAGATCTGATACC 1957
QY 1441 CATTTAGAGCCCGAGAAATTCCTCTCTCTGGATCCAGTTTGGAGCAAAACCCACACCCAG 1500
DB 1958 CATTTAGAGCCCGAGAAATTCCTCTCTCTGGATCCAGTTTGGAGCAAAACCCACACCCAG 2017
QY 1501 CTCACACACAAACAAATGGACAGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGCG 1560
DB 2018 CTCACACACAAACAAATGGACAGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGCG 2077
QY 1561 TGTGTTGGAGCCTCCCGAGTAACACCTATTTATTTTACCTCTTTCCCAAACTGGAGCA 1620
DB 2078 TGTGTTGGAGCCTCCCGAGTAACACCTATTTATTTTACCTCTTTCCCAAACTGGAGCA 2137
QY 1621 TTTATGCTTAGCTGTCAAGATCTGTCAGTCCCTCTCCTTCTCAATAAAGCATCTT 1680
DB 2138 TTTATGCTTAGCTGTCAAGATCTGTCAGTCCCTCTCCTTCTCAATAAAGCATCTT 2197
QY 1681 CAAGCTTGTAAAAAATAA 1703
DB 2198 CAAGCTTGTAAAAAATAA 2220
RESULT 3
ID ABA91311
XX ABA91311 standard; DNA; 2159 BP.
AC ABA91311;
DT 08-APR-2002 (first entry)
DE Human colon specific gene 337950.
XX Colon specific gene; CSG; gene; human; colon cancer; tumour;
KW metastasis; diagnosis; gene therapy; imaging; staging; vaccine; ds.
XX Homo sapiens.
XX WO200192528-A2.
XX 06-DEC-2001.
XX 29-MAY-2001; 2001WO-US17583.
XX 26-MAY-2000; 2000US-207383P.
XX (DIAD-) DIADEXUS INC.
PA Macina RA, Chen S, Pluta J, Sun Y, Recipon H;
XX WPI; 2002-083188/11.
XX Colon specific polypeptides and polynucleotides useful for diagnosing,
PT monitoring, staging and treating colon cancer -
XX Claim 1; Page 101-102; 116pp; English.
XX The present sequence is that of a colon specific gene (CSG),
CC designated LSGold gene 337950. This is 1 of 22 claimed CSGs (see

CC ABA91302-23) identified in the Incyte LIFESEQ Gold database using
CC the data mining Cancer Leads Automatic Search Package (CLASP),
CC which identifies highly expressed organ and cancer specific genes.
CC The CSG, or its encoded protein, can be used as a diagnostic marker
CC of colon cancer. The invention provides claimed methods of
CC diagnosing colon cancer, diagnosing metastasis of colon cancer,
CC metastasis, and monitoring a change in stage of colon cancer in a
CC patient, by determining the level of a CSG in a sample of cells,
CC tissues or body fluid and comparing it with the level in a sample
CC corresponding sample from a healthy human control or from a sample
CC taken previously from the patient. Also claimed are methods
CC of identifying therapeutic agents for use in imaging and treating
CC colon cancer. Treatment may involve administering a compound
CC which downregulates expression or activity of a CSG. Adoptive
CC immunotherapy and gene therapy methods may be used. A claimed
CC vaccine for treating colon cancer comprises a claimed CSG.
XX
SQ Sequence 2159 BP; 528 A; 673 C; 581 G; 377 T; 0 other;
Query Match 78.0%; Score 1334.6; DB 24; Length 2159;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1580; Conservative 0; Mismatches 29; Indels 108; Gaps 12;
QY 1 GGCAGAGGAGCTGAAGACAGGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC 60
DB 532 GGCAGAGGAGCTGAAGACAGGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC 591
QY 61 TCGACTTTGGAGGCTTCAGCCAGGCGGACAGATGGAGGGGCTCTGTATGGAAGGCC 120
DB 592 TCGACTTTGGAGGCTTCAGCCAGGCGGACAGATGGAGGGGCTCTGTATGGAAGGCC 651
QY 121 GCTCCTATGGAGGAGGACGCTATCTGGAGCGGGATCCCTCCAGAACAGCCACCA 180
DB 652 GCTCCTATGGAGGAGGACGCTATCTGGAGCGGGATCCCTCCAGAACAGCCACCA 711
QY 181 GAGGACCTTAGAGCAGGCTCCACCATCCCAAGGCGCTCCAGCCACACACAGTGC 240
DB 712 GAGGACCTTAGAGCAGGCTCCACCATCCCAAGGCGCTCCAGCCACACAGTGC 771
QY 241 CCGAAGACCAAGTGCCTTTACTCTGCTCCCAAGGCGCTCTTCCCGGAGACCC 300
DB 772 CCGAAGACCAAGTGCCTTTACTCTGCTCCCAAGGCGCTCTTCCCGGAGACCC 831
QY 301 AGAGAGGAGGAGGAGTGTCTCAACATGTCTTAAGGACATTTAGCTGTTCATGGAAA 360
DB 832 AGAGAGGAGGAGGAGTGTCTCAACATGTCTTAAGGACATTTAGCTGTTCATGGAAA 891
QY 361 GCTGGAGAGGCCCGCCAGCAAAACAGCAGGAGGAAGAAATTTGGGAAAAA---AACAA 417
DB 892 GCTGGAGAGGCCCGCCAGCAAAACAGCAGGAGGAAGAAATTTGGGAAAAAAGACA 951
QY 418 GGACAGGAGGAGTCTCACCCAGGACAGTACA-TTGTACTGTCTTCAGAGATCAAGTACA 476
DB 952 GGACAGGAGGAGTCTCACCCAGGACAGTACA-TTGTACTGTCTTCAGAGATCAAGTACA 1011
QY 477 GCTTCAACCTCTCTGGAGGCTGGCCACCTGGCTCAAGGAGACAGTGCCTCTGAGCTCG 536
DB 1012 GCTTCAACCTCTCTGGAGGCTGGCCACCTGGCTCAAGGAGACAGTGCCTCTGAGCTCG 1071
QY 537 TACACATCTCTTCAAGTCCCTTGAATTCATCTGCGCCAGGTGCCCTGAGGCTGCGCTAG 596
DB 1072 TACACATCTCTTCAAGTCCCTTGAATTCATCTGCGCCAGGTGCCCTGAGGCTGCGCTAG 1131
QY 597 CAGCCCAAGTGTATCTACCCCTTCTCCACCTTAAAGCTATCAACCTGTACAGTCTCTGTC 656
DB 1132 CAGCCCAAGTGTATCTACCCCTTCTCCACCTTAAAGCTATCAACCTGTACAGTCTCTGTC 1191
QY 657 TAAGCCCACTGAGAGTAACTTTTGGATGGGTTTGGGCCAGCCTGGACACTAGCCGGG 716
DB 1192 TAAGCTACCTGAGAGTAACTTTTGGATGGGTTTGGGCCAGCCTGGACACTAGCCGGG 1251
QY 717 CCGACTGGACAGGCGATGAGCCCTGCCCTACCAACCCACATTTCTCAGATGACTGGCAAC 776

Db 1252 CCGACTGGACAGGGGATGAGCCCTGAGCCCTACCAACCCACATTTCTCAGATGACTGGCAAC 1311
Qy 777 TTCAGAGCCCTCCAGCCAAGCACCCTTAGGATACAGGACCCCTGTTTCCTTCGCGGG 836
Db 1312 TTCAGAGCCCTCCAGCCAAGCACCCTTAGGATACAGGACCCCTGTTTCCTTCG----- 1366
Qy 837 GAAGTCATAGGTTAGGGAGCACCTTCACACTTCTCTCAGGAGAAGACACACAACCATGACC 896
Db 1367 ----- 1366
Qy 897 CTCAGCCTGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCACCCAG-CCCTG 955
Db 1367 -----GGCCTCCAGTCCCAACCTGCCACCCAGTCCCTG 1402
Qy 956 AAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAA--CCACAGGGAACCTCACTGTGTGTC 1013
Db 1403 AAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAAATCCACAGGGAACCTCACTGTGTGTC 1462
Qy 1014 AGGAGAGAAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGTGAAGAAATGAGG 1073
Db 1463 AGGTAGAGAAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGGTGAAGAAATGAGG 1522
Qy 1074 CGGACCGAGCGGCTACATTTCCAAAGCAACATCTCTGGAGCCCTACAGCCGGGGAACCCCTG 1133
Db 1523 CGGACCGAGCGGCTACATTTCCAAAGCAACATCTCTGGAGCCCTACAGCCGGGGAACCCCTG 1582
Qy 1134 GGACCCAGGCGGCTAGTCA-CCCTCTCGGGTTCATGCTTCGACTTAGCTCGAGGCCGTGAA 1192
Db 1583 GGACCCAGGCGGCTAGTCAACCCCTCTCGGGTTCATGCTTCGACTTAGCTCGAGGCCGTGAA 1642
Qy 1193 GAGTCAACAGCTGGCTGTCAGGAGCAAACTTCTCCACTCCAGCGGTGAGCACTTGGG 1252
Db 1643 GAGTCAACAGCTGGCTGTCAGGAGCAAACTTCTCCACTCCAGCGGTGAGCACTTGGG 1702
Qy 1253 TCCCTGAGGGG--AGCCAGCTACTTGGCA-TAAGACTCGGGAGCTAC--AGATGCTAT 1307
Db 1703 TCCCTGAGGGGAGCCGACGCTACTTGGCAATTAAGACTCGGGAGCTTACGAGTATGCTAT 1762
Qy 1308 GTCCACAGGAGGCCACAC---GAATCTGTCCGGCTGAGGCTGTGAGAGGATGC-TG 1363
Db 1763 GTCCACAGGAGGCCACACGAAATCTGTCCGGCTGAGGCTGTGAGAGGATGCTG 1822
Qy 1364 GGGATAAGCCCTTAGGCACCAAGCTTAGACACCTCCAAAGAACCGGCCCTGATGCAAG 1423
Db 1823 GGGATAAGCCCTTAGGCACCAAGCTTAGACACCTCCAAAGAACCGGCCCTGATGCAAG 1882
Qy 1424 ATGCGAGATCTGATACCCATAGAGCCCGGAGAAATTCCTCTTCTGGATCCAGTTTGCAG 1483
Db 1883 ATGCGAGATCTGATACCCATAGAGCCCGGAGAAATTCCTCTTCTGGATCCAGTTTGCAG 1942
Qy 1484 CAACCCCAACAC--CCGAGCTCACACAGCAAAACATGGACAGGCCCGAGGCTGAAGC 1541
Db 1943 CAACCCCAACACCTCCAGGCTCACACAGCAAAACATGGACAGGCCCGAGGCTGAAGC 2002
Qy 1542 AAACAGTGTCCCTTCTGGCTGTGTGGAGCTCCCAAGTAACCACTATTTATTTACCT 1601
Db 2003 AAACAGTGTCCCTTCTGGCTGTGTGGAGCTTCCCAAGTAACCACTATTTATTTACCT 2062
Qy 1602 CTTTCCCAACCTGGAGCATTTATGCTTAGCTAGGCTGTGCAAGAAATCTGTTAGTCCCTCTCC 1661
Db 2063 CTTTCCCAACCTGGAGCATTTATGCTTAGGCTGTGCAAGAAATCTGTTAGTCCCTCTCC 2122
Qy 1662 TTCTCAATAAAGCATCTTCAAGCTTCTAAAAA 1698
Db 2123 TTCTCAATAAAGCATCTTCAAGCTTGA 2159
RESULT 4
ABK43674
ID ABK43674 standard; cdna; 893 BP.
XX
AC ABK43674;

XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #254.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0236803.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-581633/65.
XX P-PSDB; AAU87344.
XX
XX New isolated nucleic acid encoding a protein for diagnosing,
XX preventing, treating or ameliorating medical conditions and used as
XX food additives or preservatives -
XX
XX Claim 1; SEQ ID NO 264; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
XX novel central nervous system protein. (I) and polypeptides (III) encoded
XX by (I), are used to treat a medical conditions and in diagnosis of a
XX pathological condition. Disorders which are diagnosed or treated include
XX autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angiogenesis, nervous system disorders e.g. Alzheimer's disease and
XX amyotrophic lateral sclerosis, infections caused by bacteria, viruses
XX e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
XX e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
XX adenocarcinomas and irritable bowel syndrome, reproductive system
XX disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
XX and pituitary dwarfism, cancers and disorders at the cellular level e.g.
XX leukaemia, disorders involving neovascularisation e.g. malignancies,
XX respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
XX acute kidney failure and blood related disorders e.g. myocardial
XX infarction. The polypeptides can also be used to aid wound healing and
XX epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX maintain organs before transplantation, for supporting cell culture of
XX primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities, fat content, lipid, protein,
XX
Query Match 48.7%; Score 832.8; DB 23; Length 893;
Best Local Similarity 99.1%; Pred. No. 9.2e-231;
Matches 837; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 866 TTTCTCAGGAGAGACACACACATGACCTCAGCTGGGGACCCCACTCCAGGCC 925
Db 48 TTCCCTCGGAGAGACACACACATGACCTCAGCTGGGGACCCCACTCCAGGCC 107
Qy 926 TCCAGCCCCAACCTGCCAGCCAGCCCTGAATGCAAGTCTTGTAGGAGTTGAAGCT 985
Db 108 TCCAGCCCCAACCTGCCAGCCAGCCCTGAATGCAAGTCTTGTAGGAGTTGAAGCT 167
Qy 986 AGGAACCCACGGGAACCTGACTGTGTCAGGAGAGAGCTGGAGGTCTCGACACAGC 1045
Db 168 AGGAACCCACGGGAACCTGACTGTGTCAGGAGAGAGCTGGAGGTCTCGACACAGC 227
Qy 1046 AAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGCGGCTACATTCCAAAGCAATC 1105
Db 228 AAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGCGGCTACATTCCAAAGCAATC 287
Qy 1106 CTGAGGCCCTTACAGCCGGGGACCCCTCGGGACCCAGGCCAGTCACCTCTCGGGTTCCA 1165

|||||
Db 288 CTGGAGCCCTACAGCGGGGACCCCTGGGACCCAGGSCCAGTCAACCTCTCGGGTTCCA 347
Qy 1166 ATGCTTCGACTTACTCGAGCCCTGAAGAGGTCACAGACTGGCTGCAGGAGAGAACTTC 1225
Db 348 ATGCTTCGACTTACTCGAGCCCTGAAGAGGTCACAGACTGGCTGCAGGAGAGAACTTC 407
Qy 1226 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 1285
Db 408 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 467
Qy 1286 CCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACGAATTCCTGTCCCGGCTGGAG 1345
Db 468 CCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACGAATTCCTGTCCCGGCTGGAG 527
Qy 1346 GCTGTCAGAGGATGCTGGGATAGCCCTTAGGCACACGCTTAGACACCTCCAGAAC 1405
Db 528 GCTGTCAGAGGATGCTGGGATAGCCCTTAGGCACACGCTTAGACACCTCCAGAAC 587
Qy 1406 AGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCGAGAAATTCCTCTT 1465
Db 588 AGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCGAGAAATTCCTCTT 647
Qy 1466 CTGATCCAGTTTGCAGCAAAACCCACAGCCAGCTCACAGCAAAACAAATGGACAG 1525
Db 648 CTGATCCAGTTTGCAGCAAAACCCACAGCCAGCTCACAGCAAAACAAATGGACAG 707
Qy 1526 GCCCAGAGGCTGAAGCAACAGTCTCCCTCTGGCTGTGTGGAGCCCTCCCAAGTACCA 1585
Db 708 GCCCAGAGGCTGAAGCAACAGTCTCCCTCTGGCTGTGTGGAGCCCTCCCAAGTACCA 767
Qy 1586 CCTATTATTTTACCTCTTTTCCCAACCTGGAGCATTTATGCCTTAGGCTTGTCAAGAAATC 1645
Db 768 CCTATTATTTTACCTCTTTTCCCAACCTGGAGCATTTATGCCTTAGGCTTGTCAAGAAATC 827
Qy 1646 TGTTCAGTCCCTCTCTCTCAATAAAGACATCTTCAAGCTTGTAAATAAATAAANG 1705
Db 828 TGTTCAGTCCCTCTCTCTCAATAAAGACATCTTCAAGCTTGTAAATAAATAAANG 887
Qy 1706 ATAAA 1710
Db 888 AAAAA 892
RESULT 5
AAH33515
ID AAH33515 standard; cDNA; 872 BP.
AC AAH33515;
XX 03-SEP-2001 (first entry)
DT Human colon cancer antigen encoding cDNA SEQ ID NO:571.
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX Homo sapiens.
OS
XX WO200122920-A2.
PN
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR

DR P-PSDB; AAG74084.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
XX useful for preventing, diagnosing and/or treating colorectal cancers -
PS Claim 1; Page 2629; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 872 BP; 225 A; 261 C; 218 G; 164 T; 4 other;

Query Match 48.6%; Score 830.8; DB 22; Length 872;
Best Local Similarity 99.4%; Pred. No. 3.5e-230;
Matches 832; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 866 TTTCTCTCAGAGAGACACACACACCATGACCCTCAGCTGGGGACCCCACTCCAGGCC 925
Db 30 TTCCCTTCGGAGAGACACACACCATGACCCTCAGCTGGGGACCCCACTCCAGGCC 89
Qy 926 TCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCT 985
Db 90 TCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCT 149
Qy 986 AGGAACCCAGGGAACTGACTGTGGTCCAGGGAGAGAGCTGGAGGTTCTGGACACAGC 1045
Db 150 AGGAACCCAGGGAACTGACTGTGGTCCAGGGAGAGAGCTGGAGGTTCTGGACACAGC 209
Qy 1046 AAGCGTGTGGCTGTGAAGATGAGCGGGAGCGGCTACATTTCCAAGCAACATC 1105
Db 210 AAGCGTGTGGCTGTGAAGATGAGCGGGAGCGGCTACATTTCCAAGCAACATC 269
Qy 1106 CTGGAGCCCTTACAGCGGGGACCCCTGGGACCCAGGSCCAGTCACTCTCGGGTTCCA 1165
Db 270 CTGGAGCCCTTACAGCGGGGACCCCTGGGACCCAGGSCCAGTCACTCTCGGGTTCCA 329
Qy 1166 ATGCTTCGACTTAGCTCAGGCTTGAAGAGTTCACAGACTGGCTGCAGGAGAGAACTTC 1225
Db 330 ATGCTTCGACTTAGCTCAGGCTTGAAGAGTTCACAGACTGGCTGCAGGAGAGAACTTC 389
Qy 1226 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 1285
Db 390 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 449
Qy 1286 CTGGGAGGCTTACAGATGCTATGTCCACAGGAGGCCCCACGAATCTGTCCCGGCTGGAG 1345
Db 450 CTGGGAGGCTTACAGATGCTATGTCCACAGGAGGCCCCACGAATCTGTCCCGGCTGGAG 509
Qy 1346 GCTGTCAGAAGGATGCTGGGATAGCCCTTAGGCACACGCTTAGACACCTCCAGAAC 1405
Db 510 GCTGTCAGAAGGATGCTGGGATAGCCCTTAGGCACACGCTTAGACACCTCCAGAAC 569
Qy 1406 AGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCGAGAAATTCCTCTT 1465
Db 570 AGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCGAGAAATTCCTCTT 629

QY 1466 CTGGATCCAGTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGGACAG 1525
Db 630 CTGGATCCAGTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGGACAG 689
QY 1526 GCCCAGAGGCTGAAGCAACAGTGCCTTCTGGCTGTGTTGGAGCCTCCCGAGTAACCA 1585
Db 690 GCCCAGAGGCTGAAGCAACAGTGCCTTCTGGCTGTGTTGGAGCCTCCCGAGTAACCA 749
QY 1586 COTATTTATTTACCTCTTCCCAAACTGGAGCAATTTATGCTAGGCTTGTCAAGAATC 1645
Db 750 COTATTTATTTACCTCTTCCCAAACTGGAGCAATTTATGCTAGGCTTGTCAAGAATC 809
QY 1646 TGTTCAGTCCCTCTCTCTCAATAAAAGCATCTTCAAGCTGTGTAATAAAAAATA 1702
Db 810 TGTTCAGTCCCTCTCTCTCAATAAAAGCATCTTCAAGCTGTGTAATAAAAAANA 866

RESULT 6
AAS27109
ID AAS27109 standard; cDNA; 872 BP.
XX
AC AAS27109;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seq ID 144.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antineoplastic; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
PN WO200154733-Al.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

14-AUG-2000; 2000US-0225757.
14-AUG-2000; 2000US-0225758.
14-AUG-2000; 2000US-0225759.
18-AUG-2000; 2000US-0226279.
22-AUG-2000; 2000US-0226681.
22-AUG-2000; 2000US-0226868.
22-AUG-2000; 2000US-0227182.
23-AUG-2000; 2000US-0227009.
30-AUG-2000; 2000US-0228924.
01-SEP-2000; 2000US-0229287.
01-SEP-2000; 2000US-0229343.
01-SEP-2000; 2000US-0229344.
01-SEP-2000; 2000US-0229345.
05-SEP-2000; 2000US-0229509.
05-SEP-2000; 2000US-0229513.
06-SEP-2000; 2000US-0230437.
06-SEP-2000; 2000US-0230438.
08-SEP-2000; 2000US-0231242.
08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
02-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246533.

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; anglogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
OS Homo sapiens.
XX WO200155318-A2.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01332.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226581.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

Qy	1226	TCCACTGCCACGGTGGAGGACACTTGGGTCCCTGACGGGGAGCGAGCTACTTCGCATAAGA	1285
Db	390	TCCACTGCCACGGTGGAGGACACTTGGGTCCCTGACGGGGAGCGAGCTACTTCGCATAAGA	449
Qy	1286	CTGTGGGAGGTACAGATGCTATGTCACAGGAGGGCCCCACGAAATCCTGTCCCGGCTGGAG	1345
Db	450	CCTGGGAGGCTACAGATGCTATGTCACAGGAGGGCCCCACGAAATCCTGTCCCGGCTGGAG	509
Qy	1346	GCCTCTCAGAGGAGTCTGGGGATAAGCCCTTAGGGACACAGCTTAGACACCTTCCAAAGAACC	1405
Db	510	GCCTCTCAGAGGAGTCTGGGGATAAGCCCTTAGGGACACAGCTTAGACACCTTCCAAAGAACC	569
Qy	1406	AGGCCCCGCTGATGCAAGATGGCGAGATCTGATACCCATTAGAGCCCGCAGAAATTCCTCTT	1465
Db	570	AGGCCCCGCTGATGCAAGATGGCGAGATCTGATACCCATTAGAGCCCGCAGAAATTCCTCTT	629
Qy	1466	CTGGATCCCCAGTTTGGCAGCAAAACCCACACCCAGCTCACACGAGCAAAACAAATGGACAG	1525
Db	630	CTGGATCCCCAGTTTGGCAGCAAAACCCACACCCAGCTCACACGAGCAAAACAAATGGACAG	689
Qy	1526	GCCGAGAGGCTGAAGCAACAGTGTCCCTTCTGGCTGTGTTGGAGGCTCCCCAGTTAAACA	1585
Db	690	GCCGAGAGGCTGAAGCAACAGTGTCCCTTCTGGCTGTGTTGGAGGCTCCCCAGTTAAACA	749
Qy	1586	CTATATTTATTTTACCTCTTCCCAAACCTGGAGCATTTATGCCTTAGGCTTGTCAAGAATC	1645
Db	750	CTATATTTATTTTACCTCTTCCCAAACCTGGAGCATTTATGCCTTAGGCTTGTCAAGAATC	809
Qy	1646	TGTTTCAGTCCCTCTCCTTCTCAATAAAAGCATCTTCAAGCTTGTAAAAAAAATA	1702
Db	810	TGTTTCAGTCCCTCTCCTTCTCAATAAAAGCATCTTCAAGCTTGTAAAAAAAATA	866
RESULT 8			
AAS27108/c			
ID AAS27108 standard; cdNA; 1265 BP.			
XX	AC	AAS27108;	
XX	DT	07-NOV-2001 (first entry)	
XX	XX		
DE	DE	cdNA encoding novel signal transduction pathway protein, Seq ID 143.	
KW	KW	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour	
KW	KW	antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;	
KW	KW	immune system disorder; rheumatoid arthritis; inflammatory condition;	
KW	KW	organ transplant rejection; infection; hepatitis C; blood disorder;	
KW	KW	sickle cell anemia; hyperproliferative disorder; Gaucher's disease;	
KW	KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;	
KW	KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;	
KW	KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease;	
KW	KW	reproductive system; gastrointestinal; liver disorder; AIDS; ss;	
OS	OS	acquired immune deficiency syndrome.	
XX	XX	Homio sapiens.	
XX	PN	WO200154733-A1.	
XX	XX	02-AUG-2001.	
XX	XX	17-JAN-2001; 2001WO-US01312.	
PR	XX	31-JAN-2000; 2000US-0179065.	
PR	PR	04-FEB-2000; 2000US-0180628.	
PR	PR	24-FEB-2000; 2000US-0184664.	
PR	PR	02-MAR-2000; 2000US-0186350.	
PR	PR	16-MAR-2000; 2000US-0189874.	
PR	PR	17-MAR-2000; 2000US-0190076.	
PR	PR	18-APR-2000; 2000US-0198123.	
PR	PR	19-MAY-2000; 2000US-0205515.	
PR	PR	07-JUN-2000; 2000US-0209467.	
PR	PR	28-JUN-2000; 2000US-0214886.	

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 13-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 20-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465460/50.
XX P-PSDB; AAU17191.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
XX prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders
XX
XX Claim 1; SEQ ID No 143; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
XX polynucleotides (II). (I), (II) and the antibody to (I) are useful for
XX diagnosing, preventing and treating diseases including immune system
XX disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
XX disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
XX transplant rejections and graft versus host disease, infectious diseases

RESULT 9
ABK43673

PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2000US-0254097.
PR 05-JAN-2001; 2000US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPT; 2001-581633/65.
XX P-PSDB; AAU87343.
DR
DR
XX
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
XX
PS Claim 1; SEQ ID No 263; 837pp; English.
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
Query Match 39.3%; Score 671.8; DB 23; Length 1282;
Best Local Similarity 99.0%; Pred. No. 5.2e-184;
Matches 673; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 1027 GGAGGTTCTGGACCACAGCGGTGGTGGTGGTGAAGATGAGCGGGACGAGCGG 1086
Db 603 GCAGGTTCTGGACCACAGCGGTGGTGGTGGTGAAGATGAGCGGGACGAGCGG 662
Qy 1087 CTACATTCCAAGCAACATCTCTGGAGCCCTACAGCGGGGACCCCTGGAGGCCA 1146
Db 563 CTACATTCCAAGCAACATCTCTGGAGCCCTACAGCGGGGACCCCTGGAGGCCA 722
Qy 1147 GTCACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCACAGCTG 1206
Db 723 GTCACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCACAGCTG 782
Qy 1207 GCTGAGCAGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGGTCCTGACGGGAG 1266

Db 783 GCTGACGGCAGAGAACTTCCACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGAG 842
QY 1267 CCAGCTACTTCCGATAAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACG 1326
Db 843 CCAGCTACTTCCGATAAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACG 902
QY 1327 AATCCTGTCCCGCTGGAGGCTGTCAAGAGATGCTGGGATAAGCCCTTAGGCACCCAGC 1386
Db 903 AATCCTGTCCCGCTGGAGGCTGTCAAGAGATGCTGGGATAAGCCCTTAGGCACCCAGC 962
QY 1387 TTAGACACCTCCAGAACACCGCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAG 1446
Db 963 TTAGACACCTCCAGAACACCGCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAG 1022
QY 1447 AGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCGACGAAACCCACACCCAGCTCACA 1506
Db 1023 AGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCGACGAAACCCACACCCAGCTCACA 1082
QY 1507 CAGCAAAAACATGGACAGGCCCGAGGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTT 1566
Db 1083 CAGCAAAAACATGGACAGGCCCGAGGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTT 1142
QY 1567 GGAGCCTCCCGCAGTAACCACTATTTATTTTACCTCTTTCCCAACCTGGAGCATTTATG 1626
Db 1143 GGAGCCTCCCGCAGTAACCACTATTTATTTTACCTCTTTCCCAACCTGGAGCATTTATG 1202
QY 1627 CCTAGGCTTGTCAAGAAATCTGTTTCAGTCCCTCTCTCTCAATAAAAGCATCTTCAAGCT 1686
Db 1203 CCTAGGCTTGTCAAGAAATCTGTTTCAGTCCCTCTCTCTCAATAAAAGCATCTTCAAGCT 1262
QY 1687 TGTAATAAAAAAATAANGA 1706
Db 1263 TGTAATAAAAAAATAANGA 1282

RESULT 10
ABN93792/c
ID ABN93792 standard; DNA; 401 BP.
XX
AC ABN93792;
XX
DT 13-AUG-2002 (first entry)
XX
DE Gene #290 used to diagnose liver cancer.
KW Gene: liver cancer; ds: hepatocellular carcinoma; hepatotropic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
PF 02-OCT-2001; 2001WO-US30589.
XX
PR 02-OCT-2000; 2000US-237054P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
XX WPI; 2002-426119/45.
XX
PT Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
PS Claim 1; SEQ ID NO 290; 298pp; English.
XX
CC The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 401 BP; 83 A; 93 C; 124 G; 101 T; 0 other;
Query Match 22.7%; Score 389; DB 24; Length 401;
Best Local Similarity 99.8%; Pred. No. 2.6e-102;
Matches 400; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1259 ACGGGGAGCCAGCTACTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCCA-CAGGA 1317
Db 401 ACGGGGAGCCAGCTACTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCCACCAGGA 342
QY 1318 GGCCCCACCAAGTCTGTCCCGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 1377
Db 341 GGCCCCACCAAGTCTGTCCCGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 282
QY 1378 GGCACGAGCTTAGACACCTCCAAAGAACCCAGGCCCGCTGTATGCAAGATGGCAGATCTGAT 1437
Db 281 GGCACGAGCTTAGACACCTCCAAAGAACCCAGGCCCGCTGTATGCAAGATGGCAGATCTGAT 222
QY 1438 ACCATTAGAGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCCTTCAGCAAAACCCACACC 1497
Db 221 ACCATTAGAGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCCTTCAGCAAAACCCACACC 162
QY 1498 CAGCTCACACAGCAAAACCAATGGACAGGCCCCAGAGGCTGAAGCAACACAGTGTCCCTTCT 1557
Db 161 CAGCTCACACAGCAAAACCAATGGACAGGCCCCAGAGGCTGAAGCAACACAGTGTCCCTTCT 102
QY 1558 GGCTGTGTGGAGCCCTCCCGCAGTAACACCTATTTATTTTACCTCTTTTCCCAACCTGGA 1617
Db 101 GGCTGTGTGGAGCCCTCCCGCAGTAACACCTATTTATTTTACCTCTTTTCCCAACCTGGA 42
QY 1618 GCATTTATGCTAGGCTTGTCAAGAAATCTCTTCAGTCCCTC 1658
Db 41 GCATTTATGCTAGGCTTGTCAAGAAATCTCTTCAGTCCCTC 1

RESULT 11
AAK88048
ID AAK88048 standard; cDNA; 1388 BP.
XX
AC AAK88048;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen coding sequence SEQ ID NO: 364.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ss.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.
XX

XX PS Claim 1; SEQ ID NO 364; 986pp; English.

XX CC The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the

CC diagnosis, treatment and prevention of digestive system disorders,

CC including cancer, Meckel's diverticulum, bacterial or parasitic

CC infections, appendicitis, Hirschsprung's disease, chronic colitis or

CC ulcerative colitis. The present sequence is a cDNA encoding a digestive

CC system antigen of the invention.

XX SQ Sequence 1388 BP; 329 A; 432 C; 377 G; 238 T; 12 other;

Query Match 22.7%; Score 388.8; DB 22; Length 1388;

Best Local Similarity 97.5%; Pred. No. 5.7e-102;

Matches 422; Conservative 3; Mismatches 3; Indels 5; Gaps 3;

QY 1 GGCAGAGCGACTGAAGCACCAGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAA---AG 57

Db 615 GGCAGAGCGACTGAAGCACCAGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAAGCAG 674

QY 58 ACCTCGACTTGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGCTGCTATGGAAG 117

Db 675 ACCTCGACTTGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGCTGCTATGGAAG 734

QY 118 GCGCGTCCCTATGGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCA 177

Db 735 GCGCGTCCCTATGGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCA 794

QY 178 CCAGAGGACCTAGAGCAGCAGCTCCACCATCCCAAGGCCCTGCGAGCCACACACG 237

Db 795 CCAGAGGACCTAGAGCAGCAGCTCCACCATCCCAAGGCCCTGCGAGCCACACACG 854

QY 238 TGCCCGAGAACCAAGTGCCCTTACTTGCTCTCTCAAGGCGTCTCTTCCCGCCAGGA 297

Db 855 TGCCCGAGAACCAAGTGCCCTTACTTGCTCTCTCAAGGCGTCTCTTCCCGCCAGGA 914

QY 298 CCCAGAGGAGCAGGAGAGTGTGTAACCATGCTCTAAGGAGCATGAGCTGTTCATGG 357

Db 915 CCCAGAGGAGCAGGAGAGTGTGTAACCATGCTCTAAGGAGCATGGA-CTGNTCATGG 973

QY 358 AAAGCTGGAGAGGCCAGCCAGCAAGCAGCAGCAGGAGAGAGAAATTTGGGAAAAAACAA 417

Db 974 AAA-CTGGANAAGCCAGGCAAGCAGCAGCAGGAGAGAGAAATTTGGGAAAAAACAA 1032

QY 418 GGACAGGGAGGT 430

Db 1033 GGACAGGGAGGT 1045

RESULT 12

AAS27487

ID AAS27487 standard; cDNA; 572 BP.

XX AC AAS27487;

XX DF 07-NOV-2001 (first entry)

XX DE cDNA encoding novel signal transduction pathway protein, Seq ID 522.

XX KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;

KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;

KW immune system disorder; rheumatoid arthritis; inflammatory condition;

KW organ transplant rejection; infection; hepatitis C; blood disorder;

KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;

KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;

KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;

KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;

XX OS acquired immune deficiency syndrome.

OS Homo sapiens.

XX

PN XX WO200154733-A1.

PD XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01312.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225477.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 23-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0233397.

PR 14-SEP-2000; 2000US-0233398.

PR 14-SEP-2000; 2000US-0233399.

PR 14-SEP-2000; 2000US-0233400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235634.

PR 27-SEP-2000; 2000US-0235636.

Central nervous system; CNS; autoimmune disease; rheumatoid arthritis; hyperproliferative disorder; neoplasia; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; ischaemia; angiodysgenesis; nervous system disorder; Alzheimer's disease; AIDS; ocular disorder; acquired immunodeficiency virus; dysphasia; gastrointestinal disorder; adenocarcinoma; reproductive system disorder; testicular feminisation; endocrine disorder; diabetes; cancer; leukaemia; neovascularisation; respiratory disorder; renal disorder; kidney failure; blood disorder; myocardial infarction; wound healing; cell proliferation; skin aging; food additive; food preservative; gene therapy; gene; ss.

Matches 334; Conservative 0; Mismatches 7; Indels 0; Gaps 0;			
Qy	1366	GATAAGCCCTTAGCACCAGCTTAGACACCTTCAAGAACCCAGGCCCGCTCATGCAGAT	1425
Db	372	GATAAGCCCTTAGCACCAGCTTAGACACCTTCAAGAACCCAGGCCCGCTCATGCAGAT	431
Qy	1426	GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTTCTGGATCCCGTTTGACGCA	1485
Db	432	GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTTCTGGATCCCGTTTGACGCA	491
Qy	1486	AACCCACACCCAGCTCACACAGCAAAACATGGACGCCCGAGAGCTGAAGCAAC	1545
Db	492	AACCCACACCCAGCTCACACAGCAAAACATGGACGCCCGAGAGCTGAAGCAAC	551
Qy	1546	AGTGCCCTTCTGGCTGTGTGGAGCCTCCCGAGTACCACTATTTATTTACCTCTTT	1605
Db	552	AGTGCCCTTCTGGCTGTGTGGAGCCTCCCGAGTACCACTATTTATTTACCTCTTT	611
Qy	1606	CCCAACCTGGAGCATTTATGCCTAGCTTGTCAAGAAATCTGTTACGTCCTCTCTCTCT	1665
Db	612	CCCAACCTGGAGCATTTATGCCTAGCTTGTCAAGAAATCTGTTACGTCCTCTCTCTCT	671
Qy	1666	CAATAAAGCATCTTCAAGCTTGTAAAAAATAAANGA	1706
Db	672	CAATAAAGCATCTTCAAGCTTAAAAAATAAANGA	712
RESULT 15			
AAS27685/c			
ID	AAS27685 standard; DNA; 2643 BP.		
XX	AC AAS27685;		
XX	AC AAS27685;		
DT	07-NOV-2001 (first entry)		
XX	DNA encoding novel signal transduction pathway protein, Seq ID 1345.		
XX	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;		
KW	antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;		
KW	immune system disorder; rheumatoid arthritis; inflammatory condition;		
KW	organ transplant rejection; infection; hepatitis C; blood disorder;		
KW	sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;		
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;		
KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;		
KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease;		
KW	reproductive system; gastrointestinal; liver disorder; AIDS; ds;		
KW	acquired immune deficiency syndrome.		
OS	Homo sapiens.		
XX	WO200154733-A1.		
PN	02-AUG-2001.		
XX	17-JAN-2001; 2001WO-US01312.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226868.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		
PR	08-SEP-2000; 2000US-0232080.		
PR	08-SEP-2000; 2000US-0232081.		
PR	12-SEP-2000; 2000US-0231968.		
PR	14-SEP-2000; 2000US-0232397.		
PR	14-SEP-2000; 2000US-0232398.		
PR	14-SEP-2000; 2000US-0232399.		
PR	14-SEP-2000; 2000US-0232400.		
PR	14-SEP-2000; 2000US-0232401.		
PR	14-SEP-2000; 2000US-0233063.		
PR	14-SEP-2000; 2000US-0233064.		
PR	14-SEP-2000; 2000US-0233065.		
PR	21-SEP-2000; 2000US-0234223.		
PR	21-SEP-2000; 2000US-0234274.		
PR	25-SEP-2000; 2000US-0234997.		
PR	25-SEP-2000; 2000US-0234998.		
PR	26-SEP-2000; 2000US-0235484.		
PR	27-SEP-2000; 2000US-0235834.		
PR	27-SEP-2000; 2000US-0235836.		
PR	29-SEP-2000; 2000US-0236327.		
PR	29-SEP-2000; 2000US-0236367.		
PR	29-SEP-2000; 2000US-0236368.		
PR	29-SEP-2000; 2000US-0236369.		
PR	29-SEP-2000; 2000US-0236370.		
PR	02-OCT-2000; 2000US-0236802.		
PR	02-OCT-2000; 2000US-0237037.		
PR	02-OCT-2000; 2000US-0237038.		
PR	02-OCT-2000; 2000US-0237039.		
PR	02-OCT-2000; 2000US-0237040.		
PR	13-OCT-2000; 2000US-0239935.		
PR	13-OCT-2000; 2000US-0239937.		
PR	20-OCT-2000; 2000US-0240960.		
PR	20-OCT-2000; 2000US-0241221.		
PR	20-OCT-2000; 2000US-0241785.		
PR	20-OCT-2000; 2000US-0241786.		
PR	20-OCT-2000; 2000US-0241787.		
PR	20-OCT-2000; 2000US-0241808.		
PR	20-OCT-2000; 2000US-0241809.		
PR	20-OCT-2000; 2000US-0241826.		
PR	01-NOV-2000; 2000US-0244617.		
PR	08-NOV-2000; 2000US-0246474.		
PR	08-NOV-2000; 2000US-0246475.		

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249219.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 1345; 880pp; English.
XX

CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (i), (ii) and the antibody to (i) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in

CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX
SQ Sequence 2643 BP; 562 A; 729 C; 705 G; 647 T; 0 other;

Query Match 18.9%; Score 324; DB 22; Length 2643;
Best Local Similarity 100.0%; Pred. No. 4.8e-83;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1366 GATAAGCCCTTAGCACCAGCTTAGACACCTCCCAAGAACACAGGCCGCCCTGATGCAAGAT 1425
Db 324 GATAAGCCCTTAGCACCAGCTTAGACACCTCCCAAGAACACAGGCCGCCCTGATGCAAGAT 265
Qy 1426 GGCAGATCTGATACCCCATTTAGAGCCCGGAGAAATTCCTCTTCTGGATCCAGTTTGCAGCA 1485
Db 264 GGCAGATCTGATACCCCATTTAGAGCCCGGAGAAATTCCTCTTCTGGATCCAGTTTGCAGCA 205
Qy 1486 AACCCACACCCAGCTCACACAGCAAAACAATGGACAGGCCCCAGAGGCTGAAGCAAAAC 1545
Db 204 AACCCACACCCAGCTCACACAGCAAAACAATGGACAGGCCCCAGAGGCTGAAGCAAAAC 145
Qy 1546 AGTGTCCTTCTGCTGTGTGGAGCCCTCCCAAGTAACACCTATTTATTTTACTCTTT 1605
Db 144 AGTGTCCTTCTGCTGTGTGGAGCCCTCCCAAGTAACACCTATTTATTTTACTCTTT 85
Qy 1606 CCCAAACCTGGAGGATTTATGCTTAGGCTTGTCAAGAAATCTGTTCAGTCCCTCTCTTCT 1665
Db 84 CCCAAACCTGGAGGATTTATGCTTAGGCTTGTCAAGAAATCTGTTCAGTCCCTCTCTTCT 25
Qy 1666 CAATAAAAGCATCTTCAAGCTTGT 1689
Db 24 CAATAAAAGCATCTTCAAGCTTGT 1

Search completed: February 25, 2003, 07:43:06
Job time : 284 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	975.8	57.1	1651	4	US-09-484-970B-48	Sequence 48, Appl
2	155.6	9.1	197	1	US-08-700-575-40	Sequence 40, Appl
3	93.2	5.5	1635	4	US-09-484-970B-51	Sequence 51, Appl
4	49.8	2.9	3245	1	US-07-935-311A-3	Sequence 3, Appl
5	49.8	2.9	3245	1	US-08-368-079-3	Sequence 3, Appl
6	49.8	2.9	3245	5	PCR-US93-07996-3	Sequence 3, Appl
7	47.4	2.8	1457	4	US-09-444-053-3	Sequence 3, Appl
8	44.4	2.6	1803	4	US-09-006-428A-18	Sequence 18, Appl
9	44.4	2.6	1858	4	US-09-006-428A-16	Sequence 16, Appl
10	43.8	2.6	7218	1	US-08-232-463-14	Sequence 14, Appl
11	39.6	2.3	70000	4	US-09-851-896-3	Sequence 3, Appl
12	35.6	2.1	3240	4	US-09-851-896-10	Sequence 10, Appl
13	35.6	2.1	5661	4	US-08-938-105-2	Sequence 2, Appl
14	35.4	2.1	6530	2	US-08-146-930-1	Sequence 1, Appl
15	35.4	2.1	6530	1	US-08-458-240-1	Sequence 1, Appl
16	35.4	2.1	6530	5	PCR-US93-03993-1	Sequence 1, Appl
17	35.4	2.1	15894	1	US-08-348-891A-1	Sequence 1, Appl
18	35.4	2.1	15894	1	US-08-905-817-1	Sequence 1, Appl
19	35.2	2.1	16442	3	US-08-781-891-208	Sequence 208, Appl
20	34.4	2.0	4403765	4	US-09-103-840A-2	Sequence 2, Appl
21	34.4	2.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
22	34.2	2.0	480	5	PCR-US95-08295-14	Sequence 14, Appl
23	34.2	2.0	7218	1	US-08-232-463-14	Sequence 14, Appl
24	34	2.0	241	3	US-08-866-340-1	Sequence 17, Appl
25	34	2.0	4411529	4	US-09-103-840A-1	Sequence 1, Appl
26	33.8	2.0	1296	1	US-07-816-283-9	Sequence 9, Appl
27	33.8	2.0	1296	1	US-08-417-103-9	Sequence 9, Appl

```
QY 301 AGAGAGGACGAGGAGTCTGAACCATGTCTTAAGGACATTTGAGCTGTTTCATGGGAA 360
Db 820 ACAGAGGAGGAGGAGTCTGAACCATGTCTTAAGGACATTTGAGCTGTTTCATGGGAA 879
QY 361 GCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 420
Db 880 GCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 939
QY 421 CCAGGAGGAGTCTACCCAGGACAGTACATTCGACTGCTTCCAGAGATCAAGTACAGCTT 480
Db 940 CCAGGAGGAGTCTACCCAGGACAGTACATTCGACTGCTTCCAGAGATCAAGCAGCTT 999
QY 481 CAACCTCTCTGGAAGGCTGGCCACCTGGCTGAAGGAGCAAGTGGCCCTGAGCTGTACA 540
Db 1000 CAACCTCTCTGGAAGGCTGGCCACCTGGCTGAAGGAGCAAGTGGCCCTGAGCTGTACA 1059
QY 541 CATCTCTCTCAAGTCCCTGAAGTTCATCTGCTGGCCAGGTCCTGAGGCTGGCTAGCAGC 600
Db 1060 CATCTCTCTCAAGTCCCTGAAGTTCATCTGCTGGCCAGGTCCTGAGGCTGGCTAGCAGC 1119
QY 601 CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGTCTGTCTAAG 660
Db 1120 CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGTCTGTCTAAG 1179
QY 661 CCCACCTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCGGGCCGA 720
Db 1180 CTCACCTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCGGGCCGA 1239
QY 721 CTGGACAGGCGATGAGCCCTGCTTACCAACCCACATCTCAGATGACTGGCACTTCC 780
Db 1240 CTGGACAGGCGATGAGCCCTGCTTACCAACCCACATCTCAGATGACTGGCACTTCC 1299
QY 781 AGAGCCCTCCAGCAAGCACCTTAGGATACCAAGCAGCCTGTTCCCTTCGGCGGGGAG 840
Db 1300 AGAGCCCTCCAGCAAGCACCTTAGGATACCAAGCAGCCTGTTCCCTTCGGCGGGGAG 1349
QY 841 TCATAGTTAGGAGCAGCTCACACTTTCTCTCAGGAGAGACACACAACCATGACCCCTCA 900
Db 1350 ----- 1349
QY 901 GCCTGGGAGCCCAACTCCAGGCCCTCCAGCCGCCAAGCTGCCAGCCAGCCCTGAAAT 960
Db 1350 -----GBCCTCCAGCCGCCAAGCTGCCAGCCAGCCCTGAAAT 1389
QY 961 GCAAGTCTGTACGAGTTTGAAGCTAGGAACCCAGCGGAAGTGTGCTCCAGGAGGA 1020
Db 1390 GCAAGTCTGTACGAGTTTGAAGCTAGGAACCCAGCGGAAGTGTGCTCCAGGAGGA 1449
QY 1021 GAAGCTGGAGGTTCTGGACCAAGCGGTTGGGCTGGTGAAGATGAGCGGGGAGC 1080
Db 1450 GAAGCTGGAGGTTCTGGACCAAGCGGTTGGGCTGGTGAAGATGAGCGGGGAGC 1509
QY 1081 GAGCGGCTACATTCACAGCAACATCTTGAGCCCTTACAGCCGGGACCCCTGGGAGCCA 1140
Db 1510 GAGCGGCTACATTCACAGCAACATCTTGAGCCCTTAAAGCGGGGACCCCTGGGAGCCA 1569
QY 1141 GGGCCAGTCAACCTCTGGGTTTCCAGTCTGAGTCTGAGGCGCTGGAAGAGGTAC 1200
Db 1570 NGGCCAGTCAACCTCTT--GGGTTCAAGTCTGCTCG--CTTAGCTGAGGCTGGAAGGTAC 1626
QY 1201 AGACTGGCTGCAG 1213
Db 1627 AGACTGGCTGAAG 1639
```

RESULT 2
US-08-700-575-40
; Sequence 40, Application US/08700575
; Patent No. 581749
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.

```
; APPLICANT: Wilde, Craig G.  
; TITLE OF INVENTION: NOVEL HUMAN KINASE HOMOLOGS  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
; STREET: 3174 PORTER DRIVE  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/700,575  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BILLINGS, LUCY J  
; REGISTRATION NUMBER: 36749  
; REFERENCE/DOCKET NUMBER: SP-100 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 197 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; LIBRARY: Small Intestine  
; CLONE: 237002  
; US-08-700-575-40
```

```
Query Match 9.1%; Score 155.6; DB 1; Length 197;  
Best Local Similarity 96.0%; Pred. No. 1.1e-34;  
Matches 192; Conservative 0; Mismatches 4; Indels 4; Gaps 3;  
  
QY 932 CCCAAACCTGCCAGCCCTGAAATGCAAGTCTTTGTAGAGTTTGAAGCTAGGAAC 991  
Db 1 CCCAAACCTGCCAGCCCTGAAATGCAAGTCTTTGTAGAGTTTGAAGCTAGGAAC 60  
QY 992 CCACGGGAACTGACTGTGTCAGGAGAGAGCTGGAGGTTCTGGACCAAGCAAGCGG 1051  
Db 61 CCACGGGAACTGACTGTGTCAGGAGAGAGCTGGAGGTT-TGGACCAAGCAAGCGG 119  
QY 1052 TGGTGGCTGGTGAAGAAATGAGCGGCGGAGCGGCTACATTCCCAAGCAACATCTCTGGAG 1111  
Db 120 TGGTGGCTGGTGAAGAAAT-AGCGGGGAGCGGAGCTACATTCCCAAGCAACATC--TGGG 176  
QY 1112 CCCCTACAGCGGGGACCCC 1131  
Db 177 CCCCTACAGCGGGGACCCC 196
```

RESULT 3
US-09-484-970B-51
; Sequence 51, Application US/09484970B
; Patent No. 6426186
; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmuth, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program

```
; SEQ ID NO 51
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6426186 403869.2CBI
; LOCATION: 884, 1049, 1069, 1629, 1632
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-51

Query Match          5.5%; Score 93.2; DB 4; Length 1635;
Best Local Similarity 49.1%; Pred. No. 1.4e-16;
Matches 336; Conservative 0; Mismatches 338; Indels 10; Gaps 3;

QY 431 CTCACCCAGGCACATGATGCTTCCAGAAGATCAAGTACACCTTCACACCTCCTG 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 65 CCCTCGGAGGCGGAGTACCGAGTCTGCAGAAGATCAAGTACGCTTCAGCCTGCTG 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 491 GGAAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCCTGAGCTCGTACACATCCTCTTC 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 125 GCCCGGCTGGCGGCAACATCGCGACCCCTCTCTCCGGAGCTGTTGCACCTCTTTC 184
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 551 AAGTCCCTGAATTCATCTCTGGCCAGTGGCCCTGAGCTGGCCCTAGCAGCCCAAGTGATC 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 185 GGGGCTCTGCAGATGATGTGAACACCTCGGGGGGCGGAGTTCGCGAGCAGTGTGCGG 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 611 TCACCCCTCTCACCCCTAAAGCTATCAACTGCTACAGTCTCTTAAGCCCACTGAG 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 CGGCGCATCTGACATCGGATGGCGTGGCGCTGCTGCGGACAACTGCTCCACAGTGA 304
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 671 AGTAACCTTTGGAGGGTGGGCGCCAGCCTGGACCACTAGCCGGCGGCTGGACAGGC 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 AACAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCCGCCGGCTGGAGCTGTCCCG 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 731 GATGAGCCCTTGCCCTACCAACCCACATCTCTCAGATGACTGGCAACTTCCAGAGCCCTCC 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 365 GAGGAGGACCCCATACAGACCCGAGTCTTTCAGCGGCTGGGAGCGCGGTCACCTGAC 424
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 791 AGCCAACACCCCTTAGATACCAAGGACCCCTGTTTCCCTTGGCGGGGGAAGTCATAGTTA 850
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 425 CCGCAGAGCGCGC---GCCTGGGAGGACCCCACTGAGTGAAGAACAGC---TACAGCAGCAGCGG 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 851 GGGAGCACCTCACACTTCTCAGGAGGAAGACACACCAACATGACCTCAGCCTTGGGAC 910
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 479 AGGCGCGGAGCAAAAGCGCCCGAGGTCGCTGCTCAATGGTCAACGAGACTTGGGAGC 538
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 911 CCCAACTCCAGGCCCTCCAGCCCAAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTG 970
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 CAGAACTGAGCCCTCAGCTG---GAGTCAGAGACAGCAGGAAATGGGCTCTGTGTAAT 594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 971 TACAGATTTGAGCTAGAACCCACGCGGAAGTACTGCTGTGTCAGGAGAGAGAGCTGGAG 1030
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 595 TATGACTTCCAGGCCCAACAGCAGTGAAGTGTGCGTCAAGCAGCGGAGCTACTGGAG 654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1031 GTTCTGGACACACAGCAGCGGTGTGCTGGTGAAGATGAGGCGGACGAGCGGCTAC 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 655 GTCTGTGATGACATCTGTAAGTGTGAAGGTTCGGACCCAGCGGCGGAGGGAGATAT 714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1091 ATTCAGCAACATCTCTGGAGCC 1114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 715 GTGCCCTACACATCTCTGACACC 738
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 4
US-07-935-311A-3
; Sequence 3, Application US/07935311A
; Patent No. 5378809
; GENERAL INFORMATION:
; APPLICANT: Di Fiore, Pier Paolo
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: Substrate of the Epidermal Growth
```

```
; TITLE OF INVENTION: Factor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,311A
; FILING DATE: 19920825
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israel, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH035.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3245 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 246..2708
; US-07-935-311A-3

Query Match          2.9%; Score 49.8; DB 1; Length 3245;
Best Local Similarity 47.8%; Pred. No. 0.00037;
Matches 144; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 444 AGTACATTTGACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCTCGGAGGCTGGCCA 503
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1225 AGTTGTTGACTGTTTCCAGAAGTTTAAACATGGATTCAACCTTCTGGCCAAGTTGAAGT 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 504 CTTGGCTGAAGAGACAAAGTCCCTCAGCTCGTACACATCTCTTCAAGTCCCTGAAT 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1285 CCCATATCCAGAACCAGTGTTCAGATCTGGTTTCTTTTGTCTTCTCCACTAAATA 1344
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 564 TCATCTCTGGCCAGGTGCTGAGGCTGGCTTAGCAGCCCAAGTGATCTCACCCCTCCTCA 623
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1345 TGGTGGTCCAGGCAACAGTGGCCCTGAAGTGGCCAGTTGGTACTCAGCCACTGTTGA 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 624 CCCTAAAGCTATCAACCTGCTACAGTCTCTTAAGCCCACTCGAGAGTAACCTTTGGA 683
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1405 CAAAAGACACAGTTGATTTCTTAAACTACACAGCCACTGCGGAGGAAACGGAAGTGTGA 1464
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 684 TGGGTTGGCCCAAGCTGGACCACTAGCCGGCCGACTGACAGGCGATGACCCCTGC 743
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1465 TGTACTGGGAGATAGTTGGGTGAAGTGAAGAGCAGAGTGGCCGAAAGAACAGTTTCATCC 1524
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 744 C 744
Db 1525 C 1525
```

```
RESULT 5
US-08-368-079-3
; Sequence 3, Application US/08368079
; Patent No. 5610018
; GENERAL INFORMATION:
; APPLICANT: Di Fiore, Pier Paolo
; APPLICANT: Fazioli, Francesca
```

;; TITLE OF INVENTION: eps8, A Substrate for the Epidermal Growth Factor
;; TITLE OF INVENTION: Kinase, Antibodies Thereof, and Methods of Use Thereof
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: knobbe, Martens, Olson & Bear
;; STREET: 620 Newport Center Drive, 16th Floor
;; CITY: Newport Beach
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92660
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 1.5
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/368,079
;; FILING DATE: 03-JAN-1995
;; CLASSIFICATION: 530
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 07/935,311
;; FILING DATE: 25-AUG-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Israelson, Ned A.
;; REGISTRATION NUMBER: 29,655
;; REFERENCE/DOCKET NUMBER: NIH035.001DV1
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 235-8550
;; TELEFAX: (619) 235-0176
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3245 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 246..2708
US-08-368-079-3

Query Match 2.9%; Score 49.8; DB 1; Length 3245;
Best Local Similarity 47.8%; Pred. No. 0.00037;
Matches 144; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
QY 444 ACTACATTGACGTCTCCAGAGATCAAGTACAGCTTCAACCTCTCGGAGGCTGGCCA 503
Db 1225 AGTTTGTGACGTGTTCCAGAAGTTAAACATGGATTCAACCTCTGGCCAAGTTGAAGT 1284
QY 504 CTGCTGGAAGGAGACAGTGCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAAC 563
Db 1285 CCATATCCAGAACCCGAGTCTTCAGATCTGGTTCAATTTTGTACTCCACTAAATA 1344
QY 564 TCATCCTGGCCAGGTGCCCTGAGGCTGAGCCAGCCCAAGTATCTACCCCTCTCA 623
Db 1345 TGGTGTCCAGGCAACAGGTGCCCTGAACTGGCCAGTTGCGGTACTCAGCCCACTGTGA 1404
QY 624 CCCCTAAAGCTATCAACCTGCTACAGTCTCTGTCTAAGCCCACTCAGAGTAACTTGA 683
Db 1405 CAAAGACACAGTTGATTTCTTTAAACTACAGCCCACTCGCGGAGGAACGGAAGCTGTGA 1464
QY 684 TGGGTTGGGCGCCAGCTAGCCGCGGCGGAGTGGACAGGCGGATGAGCCCTGC 743
Db 1465 TGTCACTGGAGATAGTTGGGTGAAGTGAGAGCAGAGTGGCCGGAAGACAGATTCATCC 1524
QY 744 C 744
Db 1525 C 1525

RESULT 6
PCT-US93-07996-3
; Sequence 3, Application PC/TUS9307996
; t t t

;; GENERAL INFORMATION:
;; APPLICANT: The Government of the United States, as represented by the
;; APPLICANT: Secretary of Health and Human Services
;; TITLE OF INVENTION: Substrate of the Epidermal Growth Factor Kinase
;; NUMBER OF SEQUENCES: 4
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Knobbe, Martens, Olson & Bear
;; STREET: 620 Newport Center Drive, 16th Floor
;; CITY: Newport Beach
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 92660
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/07996
;; FILING DATE: 19930825
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (619) 235-8550
;; TELEFAX: (619) 235-0176
;; INFORMATION FOR SEQ ID NO: 3:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3245 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: DNA (genomic)
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 246..2708
PCT-US93-07996-3
Query Match 2.9%; Score 49.8; DB 5; Length 3245;
Best Local Similarity 47.8%; Pred. No. 0.00037;
Matches 144; Conservative 0; Mismatches 157; Indels 0; Gaps 0;
QY 444 AGTACATTGACGTCTCCAGAGATCAAGTACAGCTTCAACCTCTCGGAGGCTGGCCA 503
Db 1225 AGTTTGTGACGTGTTCCAGAAGTTTAAACATGGATTCAACCTCTGGCCAAGTTGAAGT 1284
QY 504 CTGCTGGAAGGAGACAGTGCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAAC 563
Db 1285 CCATATCCAGAACCCGAGTCTTCAGATCTGGTTCAATTTTGTACTCCACTAAATA 1344
QY 564 TCATCCTGGCCAGGTGCCCTGAGGCTGAGCCAGCCCAAGTATCTACCCCTCTCA 623
Db 1345 TGGTGTCCAGGCAACAGGTGCCCTGAACTGGCCAGTTGCGGTACTCAGCCCACTGTGA 1404
QY 624 CCCCTAAAGCTATCAACCTGCTACAGTCTCTGTCTAAGCCCACTCAGAGTAACTTGA 683
Db 1405 CAAAGACACAGTTGATTTCTTTAAACTACAGCCCACTCGCGGAGGAACGGAAGCTGTGA 1464
QY 684 TGGGTTGGGCGCCAGCTAGCCGCGGCGGAGTGGACAGGCGGATGAGCCCTGC 743
Db 1465 TGTCACTGGAGATAGTTGGGTGAAGTGAGAGCAGAGTGGCCGGAAGACAGATTCATCC 1524
QY 744 C 744
Db 1525 C 1525
RESULT 7
US-09-444-053-3
; Sequence 3, Application US/09444053A
; Patent No. 6165728
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Lex M. Cowser
; TITLE OF INVENTION: ANTISENSE MODULATION OF NCK-2 EXPRESSION
; FILE REFERENCE: RTS-0122


```

; APPLICATION NUMBER: US/08/458,240
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/146,930
; FILING DATE:
; APPLICATION NUMBER: 07/876,286
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6530 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-458-240-1

Query Match      2.1%; Score 35.4; DB 3; Length 6530;
Best Local Similarity 49.7%; Pred. No. 6.2;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 47 CTGGAGCAAAAGACCTCGACCTTGGAGGCCCTCAGCCAGCGCCAGACAGATGGAGGGGCGCT 106
   | |||| | |||| | | | |||| | | | |||| | | | |||| | | | ||||
Db 3478 CCGGAGGAACACCTCCCGAGCTCGTGGCGCGCCACCGGAATAGCCGCCGCCGCCGCG 3419

QY 107 GCTATGGAAGCCCTCCCTATGAGGAGGACGCTATCTGGAGCGGGGATCCCTCCA 166
   |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 3418 CAGCTGGATCCACCGCCACCTGAGTAGCGCGCCACAGCTAGAGCGCGCTGCTTCCA 3359

QY 167 GAACAGCCCCACCGAGGAGCCCTAGAGCACAGCCCTCCCAAGCCCTGCGCA 226
   | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 3358 CCGCGCCCGGAATAGCCCGCTCCGCGAGCTGGAGCCACCAGCCCGCCGCTTGGCG 3299

QY 227 C 227
   |
Db 3298 C 3298
```

Search completed: February 25, 2003, 16:13:39
Job time : 226 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 08:44:57 ; Search time 81 seconds

(without alignments)

11856.832 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 1710

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaataangataaa 1710

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_NA: *
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq: *
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq: *
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq: *
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq: *
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq: *
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq: *
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq: *
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq: *
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq: *
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq: *
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq: *
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq: *
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq: *
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1508.6	88.2	2148	10	US-09-998-598-2588
2	829.6	48.5	872	9	US-09-764-868-144
3	673.8	39.4	1265	9	US-09-764-868-143
4	487	28.5	499	10	US-09-998-598-188
5	411.2	24.0	447	10	US-09-998-598-2183
6	389	22.7	401	10	US-09-880-107-290
7	335.4	19.6	572	9	US-09-764-868-522
8	324	18.9	2643	9	US-09-764-868-1345
9	119.2	7.0	216	10	US-09-925-299-321
10	68.8	4.0	595	10	US-09-925-299-578
11	59	3.5	245	10	US-09-998-598-2415
12	56.4	3.3	1639	9	US-09-764-868-99
13	56.4	3.3	1661	9	US-09-764-868-512
14	55	3.2	514	10	US-09-833-790-43
15	44.4	2.6	488	10	US-09-867-701-10401
16	39.2	2.3	1466	10	US-09-925-300-531
17	39.2	2.3	23580	10	US-09-764-860-990
18	38.2	2.2	449	10	US-09-967-768A-23
19	38	2.2	1110	10	US-09-893-737-49

20	36.8	2.2	1329	9	US-10-011-588-22	Sequence 22, Appl
21	36.8	2.2	2559	9	US-10-011-588-24	Sequence 24, Appl
22	36.6	2.1	2928	9	US-09-764-868-142	Sequence 142, App
23	36.4	2.1	252	10	US-09-969-708-484	Sequence 484, App
24	36.4	2.1	252	10	US-09-880-107-949	Sequence 949, App
25	36.4	2.1	1216	10	US-09-917-800A-1453	Sequence 1453, App
26	36.2	2.1	359	9	US-09-764-868-521	Sequence 521, App
27	36	2.1	355	10	US-09-964-824A-518	Sequence 518, App
28	36	2.1	370	10	US-09-969-708-456	Sequence 456, App
29	35.8	2.1	449	10	US-09-764-847-328	Sequence 328, App
30	35.4	2.1	681	12	US-10-002-600-19	Sequence 19, Appl
31	35.4	2.1	1042	12	US-10-002-600-20	Sequence 20, Appl
32	35.4	2.1	1109	12	US-10-002-600-21	Sequence 21, Appl
33	35.4	2.1	1123	9	US-09-822-846-357	Sequence 357, App
34	35.4	2.1	4398	9	US-10-105-989-1	Sequence 1, Appl
35	35.2	2.1	108	10	US-09-919-580-710	Sequence 710, App
36	35.2	2.1	2070	9	US-09-932-367A-27	Sequence 27, Appl
37	35.2	2.1	2160	9	US-09-932-367A-9	Sequence 9, Appl
38	35.2	2.1	2190	9	US-09-932-367A-11	Sequence 11, Appl
39	35.2	2.1	2515	9	US-09-932-367A-25	Sequence 25, Appl
40	35.2	2.1	2540	9	US-09-932-367A-26	Sequence 26, Appl
41	35.2	2.1	8867	9	US-09-932-367A-22	Sequence 22, Appl
42	35	2.0	321	10	US-09-864-761-25530	Sequence 25530, A
43	35	2.0	575	10	US-09-864-761-8864	Sequence 8864, App
44	35	2.0	2451	10	US-09-771-161A-4	Sequence 4, Appl
45	35	2.0	3684	10	US-09-822-849A-76	Sequence 76, Appl

ALIGNMENTS

RESULT 1

US-09-998-598-2588

; Sequence 2588, Application US/09998598

; Patent No. US20020150922A1

; GENERAL INFORMATION:

; APPLICANT: Stolk, John A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Chenault, Ruth A.

; APPLICANT: Meagher, Madelein Joy

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.561

; CURRENT APPLICATION NUMBER: US/09/998, 598

; CURRENT FILING DATE: 2001-11-16

; NUMBER OF SEQ ID NOS: 2606

; SOFTWARE: Corixa Invention Disclosure Database

; SEQ ID NO 2588

; LENGTH: 2148

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-998-598-2588

Query Match 88.2%; Score 1508.6; DB 10; Length 2148;
Best Local Similarity 94.4%; Pred. No. 0;
Matches 1611; Conservative 0; Mismatches 5; Indels 90; Gaps 1;

Qy	1	GGCAGCGGCTGAAGACCCAGCCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC	60
Db	532	GGCAGCGGCTGAAGACCCAGCCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC	591
Qy	61	TCGACTTGGAGGCTTCAGCCAGGACAGATGAGGGGGCTCTATGGAAGGCC	120
Db	592	TCGACTTGGAGGCTTCAGCCAGGACAGATGAGGGGGCTCTATGGAAGGCC	651
Qy	121	GCTCCCTATGGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCAACCA	180
Db	652	GCTCCCTATGGAGCAGCAGCCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCAACCA	711
Qy	181	GAGGACCTTAGACACAGCTCCACCATCCCCAAGGCCCTGCGCCAGCACACCACTGC	240
Db	712	GAGGACCTTAGACACAGCTCCACCATCCCCAAGGCCCTGCGCCAGCACACCACTGC	771

QY 241 CCGAAGAACCAAGTGCCTTTACTCTCTGCCTCCTCAAGGGGGTCTCTTCCCGGAGACCC 300
Db 772 CCGAAGAACCAAGTGCCTTTACTCTGCCTCCTCAAGGGGGTCTCTTCCCGGAGACCC 831
QY 301 AGAGAGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGCGGAAA 360
Db 832 AGAGAGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGCGGAAA 891
QY 361 GCTGAGAGGCCCCAGGCAAAAGACAGCAGGAAGAAATTTGGGAAAAAAACAAGGA 420
Db 892 GCTGAGAGGCCCCAGGCAAAAGACAGCAGGAAGAAATTTGGGAAAAAAACAAGGA 951
QY 421 CCAGGGAGGTCTCACCCAGGCACAGTACATTTGACTGTCTCCAGAAATCAAGTACAGCTT 480
Db 952 CCAGGGAGGTCTCACCCAGGCACAGTACATTTGACTGTCTCCAGAAATCAAGTACAGCTT 1011
QY 481 CAACCTCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGGCCCTGAGCTAGCAGC 540
Db 1012 CAACCTCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGGCCCTGAGCTAGCAGC 1071
QY 541 CATCTCTCTCAAGTCCCTGAACCTTCATCTTGCCCAAGGTGCCCTGAGGCTGGCCCTAGCAGC 600
Db 1072 CATCTCTCTCAAGTCCCTGAACCTTCATCTTGCCCAAGGTGCCCTGAGGCTGGCCCTAGCAGC 1131
QY 601 CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTCTACAGTCTCTGTCTAAG 660
Db 1132 CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTCTACAGTCTCTGTCTAAG 1191
QY 661 CCCACTGAGAGTAACCTTTTGATGGGGTTGGGCCACGCTGGACCCTAGCCGGGCCGA 720
Db 1192 CCCACTGAGAGTAACCTTTTGATGGGGTTGGGCCACGCTGGACCCTAGCCGGGCCGA 1251
QY 721 CTGGACAGGCGATGAGCCCTTGCCTTACCACCCACATTTCTCAGATGACTGGCAACTTCC 780
Db 1252 CTGGACAGGCGATGAGCCCTTGCCTTACCACCCACATTTCTCGGATGACTGGCAACTTCC 1311
QY 781 AGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAAGGACCTGTCTTCCCTTCGGCGGGGAG 840
Db 1312 AGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAAGGACCTGTCTTCCCTTC 1361
QY 841 TCATAGTTAGGAGACCTCACACTTTCTCAGGAGAGACACACACCATGACCCTCA 900
Db 1362 ----- 1361
QY 901 GCCTGGGACCCCAACTCCAGGCCCTCAGGCCCAAACTGCCCCAGCCAGCCCTGAAAAAT 960
Db 1362 -----GGCCCTCAGCCCAAACTGCCAGCCAGCCCTGAAAAAT 1401
QY 961 GCAAGTCTTGTAAGGTTTGAAGCTAGGAACCCACGGGAATGACTGTGGTCCAGGGAGA 1020
Db 1402 GCAAGTCTTGTAAGGTTTGAAGCTAGGAACCCACGGGAATGACTGTGGTCCAGGGAGA 1461
QY 1021 GAAGCTGGAGGTTCTGGACCAGCAAGCGGTGGTGGCTGGAAGATGAGGCGGGAGC 1080
Db 1462 GAAGCTGGAGGTTCTGGACCAGCAAGCGGTGGTGGCTGGAAGATGAGGCGGGAGC 1521
QY 1081 GAGCGGTACATTTCAAGCAACATCTTGAGGCCCTTACAGCGGGGACCCCTGGGACCCA 1140
Db 1522 GAGCGGTACATTTCAAGCAACATCTTGAGGCCCTTACAGCGGGGACCCCTGGGACCCA 1581
QY 1141 GGGCCAGTCACCCCTCTCGGGTTTCCAATGCTTCGACTTACCTCGAGGCTGGAAGAGGTAC 1200
Db 1582 GGGCCAGTCACCCCTCTCGGGTTTCCAATGCTTCGACTTACCTCGAGGCTGGAAGAGGTAC 1641
QY 1201 AGACTGGCTGACGAGGAGAACTTCTCCACTGCCAGCTGAGGACACTTGGTCTCCTGAC 1260
Db 1642 AGACTGGCTGACGAGGAGAACTTCTCCACTGCCAGCTGAGGACACTTGGTCTCCTGAC 1701
QY 1261 GGGGAGCCAGCTACTTCGATTAAGACCTGGGGAGCTACAGATGCTATGTCACAGGAGGC 1320
Db 1702 GGGGAGCCAGCTACTTCGATTAAGACCTGGGGAGCTACAGATGCTATGTCACAGGAGGC 1761
QY 1321 CCCACGAATCTCTGCCGGCTGGAGGCTGTACAGAAAGGATGCTGGGGATAAGCCCTTAGGC 1380

Db 1762 CCCACGAATCTGTCCCGCTGGAGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGGC 1821
QY 1381 ACCAGCTTAGACACCTCCAAGAACAGCCCGCTGATGCAAGATGGCAGATCTGATACC 1440
Db 1822 ACCAGCTTAGACACCTCCAAGAACAGCCCGCTGATGCAAGATGGCAGATCTGATACC 1881
QY 1441 CATTAGAGCCCGGAGAAATTCCTCTCTCGATCCCAAGTTTGGAGCAAAACCCACACCCAG 1500
Db 1882 CATTAGAGCCCGGAGAAATTCCTCTCTCGATCCCAAGTTTGGAGCAAAACCCACACCCAG 1941
QY 1501 CTCACACAGCAAAACAAATGACAGGCCAGGCGTGAAGCAAAACAGTGTCCCTTCTGGC 1560
Db 1942 CTCACACAGCAAAACAAATGACAGGCCAGGCGTGAAGCAAAACAGTGTCCCTTCTGGC 2001
QY 1561 TGTGTTGAGGCTCCCAAGTAACACCTATTATTTTACCTCTTCCCAACCTGGAGCA 1620
Db 2002 TGTGTTGAGGCTCCCAAGTAACACCTATTATTTTACCTCTTCCCAACCTGGAGCA 2061
QY 1621 TTTATGCTAGGCTGTCAAGAAATCTGTTCAGTCCCTCTCTCAATAAAAGCATCTT 1680
Db 2062 TTTATGCTAGGCTGTCAAGAAATCTGTTCAGTCCCTCTCTCAATAAAAGCATCTT 2121
QY 1681 CAAGCTTGTAAAAAATAAANGA 1706
Db 2122 CAAGCTTGTAAAAAATAAANGA 2147

RESULT 2

US-09-764-868-144
; Sequence 144, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 144
; LENGTH: 872
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (865)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (867)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (870)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-144

Query Match 48.5%; Score 829.6; DB 9; Length 872;
Best Local Similarity 99.4%; Pred. No. 2.4e-220;
Matches 832; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 866 TTTCCTCAGGAGAGACACAAACCATGACCTCAGCCTGGGGAGCCCAACTCCAGGCC 925
Db 30 TTCCCTCGGAGAGACACAAACCATGACCTCAGCCTGGGGAGCCCAACTCCAGGCC 89
QY 926 TCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTGTAGAGTTTGAAGCT 985
Db 90 TCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTGTAGAGTTTGAAGCT 149
QY 986 AGGAACCCAGCGGAACCTGACTGTGGTCCAGGAGAGAGCTGGAGGTTCTGGACCACAGC 1045
Db 150 AGGAACCCAGCGGAACCTGACTGTGGTCCAGGAGAGAGCTGGAGGTTCTGGACCACAGC 209

```
QY 1046 AAGCGGTGGTGGCTGGTGAAGAATGAGCGGGACGGAGCGGTACATTTCAAGCAATC 1105
Db 210 AAGCGGTGGTGGCTGGTGAAGAATGAGCGGGACGGAGCGGTACATTTCAAGCAATC 269
QY 1106 CTGAGGCCCTFACAGCGGGGACCCCTGGGACCCAGGCGCAGTCACCTCTCGGGTTCCA 1165
Db 270 CTGAGGCCCTFACAGCGGGGACCCCTGGGACCCAGGCGCAGTCACCTCTCGGGTTCCA 329
QY 1166 ATGCTTCGACTAGCTCGAGSCCTGAAGAGGTCAAGACTGGCTGCGAGGAGAGAACTTC 1225
Db 330 ATGCTTCGACTAGCTCGAGSCCTGAAGAGGTCAAGACTGGCTGCGAGGAGAGAACTTC 389
QY 1226 TCCACTGCCACGGTAGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAGA 1285
Db 390 TCCACTGCCACGGTAGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAGA 449
QY 1286 CTTGGGAGCTFACAGATGCTATGTCACAGAGGCCGCCAGAACTTCCTCCGGCTGGAG 1345
Db 450 CTTGGGAGCTFACAGATGCTATGTCACAGAGGCCGCCAGAACTTCCTCCGGCTGGAG 509
QY 1346 GCTGTCAAGAGGATGCTGGGGATGAAGCCCTTAGGCACCCAGCTTAGACACCTTCCAAGAAC 1405
Db 510 GCTGTCAAGAGGATGCTGGGGATGAAGCCCTTAGGCACCCAGCTTAGACACCTTCCAAGAAC 569
QY 1406 AGGCCCCGCTGATCCAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATTCCTTT 1465
Db 570 AGGCCCCGCTGATCCAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATTCCTTT 629
QY 1466 CTGATCCCACTTTGACAGAAACCCACACCCAGCTCACACAGCAAAACAATGGACAG 1525
Db 630 CTGATCCCACTTTGACAGAAACCCACACCCAGCTCACACAGCAAAACAATGGACAG 689
QY 1526 GCCAGAGGCTGAAGCAAAACAGTCCCTTCTGGCTGTGTGGAGCCTCCCCAGTAACCA 1585
Db 690 GCCAGAGGCTGAAGCAAAACAGTCCCTTCTGGCTGTGTGGAGCCTCCCCAGTAACCA 749
QY 1586 CCTATTATTTTACCTCTTTCCCAAACTGGAGCATTTATGCCCTAGGCTTTGTCAGAAATC 1645
Db 750 CCTATTATTTTACCTCTTTCCCAAACTGGAGCATTTATGCCCTAGGCTTTGTCAGAAATC 809
QY 1646 TGTTCAGTCCCTCTCCCTCTCAATAAAGCATCTTCAAGCTTGTAATAAATAAATA 1702
Db 810 TGTTCAGTCCCTCTCCCTCTCAATAAAGCATCTTCAAGCTTGTAATAAATAAATA 866
```

RESULT 3

US-09-764-868-143/c

```
; Sequence 143, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 143
; LENGTH: 1265
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1263)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-143
```

Query Match 39.4%; Score 673.8; DB 9; Length 1265;
Best Local Similarity 98.0%; Pred. No. 5e-177;
Matches 678; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1015 GGGAGAGAAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGAAGATGAGGC 1074

```
Db 693 KGSTGACTYTTGGCAGGTTCTGACCACAGCAAGCGGTGGTGGTGAAGATGAGGC 634
QY 1075 GGGACGAGCGGCTTACATTTCAAGCAACATCTCTGAGCCCTTACAGCCGGGACCCCTGG 1134
Db 633 GGGACGAGCGGCTTACATTTCAAGCAACATCTCTGAGCCCTTACAGCCGGGACCCCTGG 574
QY 1135 GACCCAGGGCCAGTCACCTCTCGGGTTCCAATGCTTTCGACTTACGCTCGAGGCTGAAGA 1194
Db 573 GACCCAGGGCCAGTCACCTCTCGGGTTCCAATGCTTTCGACTTACGCTCGAGGCTGAAGA 514
QY 1195 GGTCAACAGACTGGCTGACGAGCAGAGAATTTCTCCACTGCCACGGTGAAGACACTTGGGTC 1254
Db 513 GGTCAACAGACTGGCTGACGAGCAGAGAATTTCTCCACTGCCACGGTGAAGACACTTGGGTC 454
QY 1255 CCTGACGGGAGCCAGCTACTTTCATTAAGACCTGGGAGCTACAGATGCTATGCTCCACA 1314
Db 453 CCTGACGGGAGCCAGCTACTTTCATTAAGACCTGGGAGCTACAGATGCTATGCTCCACA 394
QY 1315 GGAGGCCCCACGAATCCTGTCGGGCTGGAGGCTGTGCAGAGGATGCTGGGGATAAGCCC 1374
Db 393 GGAGGCCCCACGAATCCTGTCGGGCTGGAGGCTGTGCAGAGGATGCTGGGGATAAGCCC 334
QY 1375 TTAGGCACACGCTTAGACACCTTCCAAGAACGAGCCCGCTGTGCAAGATGGCAGATCT 1434
Db 333 TTAGGCACACGCTTAGACACCTTCCAAGAACGAGCCCGCTGTGCAAGATGGCAGATCT 274
QY 1435 GATACCCATTAGAGCCCGGAGAAATTCCTTCTGAGTCCAGTTTGAGCAAAACCCACA 1494
Db 273 GATACCCATTAGAGCCCGGAGAAATTCCTTCTGAGTCCAGTTTGAGCAAAACCCACA 214
QY 1495 CCCAGCTCACACAGCAAAACAATGGACAGGCCCGCAGAGGCTGAAGCAAAACAGTGTCCCT 1554
Db 213 CCCAGCTCACACAGCAAAACAATGGACAGGCCCGCAGAGGCTGAAGCAAAACAGTGTCCCT 154
QY 1555 TCTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTTATTTTACCTCTTTCCCAAACCT 1614
Db 153 TCTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTTATTTTACCTCTTTCCCAAACCT 94
QY 1615 GGAGCATTTATGCTAGGCTTGTCAGCAAAATCTGTTCAAGTCCCTCTCTCAATAAAG 1674
Db 93 GGAGCATTTATGCTAGGCTTGTCAGCAAAATCTGTTCAAGTCCCTCTCTCAATAAAG 34
QY 1675 CATCTTCAAGCTTGTAAAAAATAAANGA 1706
Db 33 CATCTTCAAGCTTGTAAAAAATAAANGA 2
```

RESULT 4

```
US-09-998-598-188
; Sequence 188, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 188
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-188
```

Query Match 28.5%; Score 487; DB 10; Length 499;
Best Local Similarity 99.8%; Pred. No. 2.1e-125;
Matches 498; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1187 CCTGAAGAGGTACAGACTGGCTGAGGACAGAGAACTTCTCCACTGCCACGGTGAGGACA 1246
Db 1 CCTGAAGAGGTACAGACTGGCTGAGGACAGAGAACTTCTCCACTGCCACGGTGAGGACA 60
QY 1247 CTTGGGTCCCTGACGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTA 1306
Db 61 CTTGGGTCCCTGACGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTA 120
QY 1307 TGTCCACAGGAGGCCCCAGCAATCTGTCCCGGCTGG-AGGCTGTCAAGAGGATCCTGGG 1365
Db 121 TGTCCACAGGAGGCCCCAGCAATCTGTCCCGGCTGGAGGCTGTCAAGAGGATCCTGGG 180
QY 1366 GATAAGCCCTTAGGCACCAAGCTTAGACACCTCCAAAGAACAGGCCGCCGCTGATGCAAGAT 1425
Db 181 GATAAGCCCTTAGGCACCAAGCTTAGACACCTCCAAAGAACAGGCCGCCGCTGATGCAAGAT 240
QY 1426 GGCAGATCTGATACCGATTTAGAGCCCGGAGAAATTCCTCTCTGAGATCCAGTTTCAGCA 1485
Db 241 GGCAGATCTGATACCGATTTAGAGCCCGGAGAAATTCCTCTCTGAGATCCAGTTTCAGCA 300
QY 1486 AACCCACACCCAGCTCACAGCAAAACAAATGAGAGGCCAGAGGCTGAAGCAAAAC 1545
Db 301 AACCCACACCCAGCTCACAGCAAAACAAATGAGAGGCCAGAGGCTGAAGCAAAAC 360
QY 1546 AGTGCCCTCTGCTGTGTGGAGCCCTCCCGAGTAACCACTATTATTTTACCTCTTT 1605
Db 361 AGTGCCCTCTGCTGTGTGGAGCCCTCCCGAGTAACCACTATTATTTTACCTCTTT 420
QY 1606 CCCAAACCTGGAGCATTTATGCTTAGCTTAGCTGTCAAGATCTGTCAGTCCCTCTCCTTCT 1665
Db 421 CCCAAACCTGGAGCATTTATGCTTAGCTTAGCTGTCAAGATCTGTCAGTCCCTCTCCTTCT 480
QY 1666 CAATAAAGCATCTTCAAG 1684
Db 481 CAATAAAGCATCTTCAAG 499

RESULT 5

US-09-998-598-2183/c
; Sequence 2183, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: STOLK, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2183
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2183

Query Match 24.0%; Score 411.2; DB 10; Length 447;
Best Local Similarity 94.5%; Pred. No. 2.2e-104;
Matches 415; Conservative 17; Mismatches 6; Indels 1; Gaps 1;

QY 1273 ACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCAGAGGAGGCCCCACGAATCCT 1332
Db 447 ACTTGCATAAAGACCTGGGGAGCTACARATGCTATGTCCAGAGGAGGCCCCACGAATCCT 388
QY 1333 GTCCCGGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCCTTAGGCACCAAGCTTAGAC 1392
Db 387 GTCCCGGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCCTTAGGCACCAAGCTTAGAC 328
QY 1393 ACCTCCAAAGAACAGGCCGCCGCTGTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCC 1452

Db 327 ACCTCCAAAGAACAGGCCGCCGCTGTGATGCAAGATGGCAGATGTGATACCCATTAGAGCCCC 268
QY 1453 GAGAAATTCCTCTCTGGATCCAGTTTGCAGCAAAACCCACACCCAGC-TCACACAGCA 1511
Db 267 GAGAAATTCCTTTTGGATCCAGTTTGCAGCAAAACCCACACCCAGCSTTCACACAGCA 208
QY 1512 AAAACAATGGACAGGCCCCAGAGGCTGAAGCAAAACAGTCTCCCTTCTGGCTGTGTGGAGC 1571
Db 207 AAAACAATGGACAGGCCCCARAGGYTGAAGCAAAACAGTCTCCCTTCTGGCTGTGTGGAGC 148
QY 1572 CTCCCACTAACCACCTATTATTTTACCTCTTTTCCCAAACTGGAGCATTTATGCCCTAG 1631
Db 147 CTCCCACTAACCACCTATTATTTTACCTCTTTTCCCAAACTGGAGCATTTATGCCCTAG 88
QY 1632 GCTTCTCAAGATCTGTTTCAGTCCCTCTCTCTCAATAAAGAGCATCTTCAAGCTTTGTA 1691
Db 87 GCTTCTCAARAATYGTTCAGTCCCTCTCTCTTTCATTAAGAGCATYTTCAAGCTTTTAA 28
QY 1692 AAAAATAAANGATAAA 1710
Db 27 AAAAATAAANGATAA 9

RESULT 6

US-09-880-107-290/c
; Sequence 290, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 290
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AAL34985
US-09-880-107-290

Query Match 22.7%; Score 389; DB 10; Length 401;
Best Local Similarity 99.8%; Pred. No. 3e-98;
Matches 400; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1259 ACGGGGACGAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCA-CAGGA 1317
Db 401 ACGGGGACGAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACAGGA 342
QY 1318 GGCCCCAGCAATCCTGTCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 1377
Db 341 GGCCCCAGCAATCCTGTCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 282
QY 1378 GGCACACCTTAGACACCTTCCAAAGAACAGGCCGCCGCTGTGATGCAAGATGGCAGATCTGAT 1437
Db 281 GGCACACCTTAGACACCTTCCAAAGAACAGGCCGCCGCTGTGATGCAAGATGGCAGATCTGAT 222
QY 1438 ACCCAATTAGAGCCCGGAGAAATTCCTTCTTGGATCCCAAGTTTGCAGCAAAACCCACACCC 1497
Db 221 ACCCAATTAGAGCCCGGAGAAATTCCTTCTTGGATCCCAAGTTTGCAGCAAAACCCACACCC 162
QY 1498 CAGCTTCACACAGCAAAACAAATGGACAGGCCCGCAGAGGCTGAAGCAAAACAGTCTCCCTTCT 1557

Db 161 CAGCTCACACGAAAAACAATGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCT 102
Qy 1558 GGTGTGTGGAGCTCCCGAGTAACACCACTATTATTTACCTCTTTCCCAAACTGGA 1617
Db 101 GGCTGTGTGGAGCTCCCGAGTAACCACTATTATTTACCTCTTTCCCAAACTGGA 42
Qy 1618 GCATTATGCTAGGCTGTGCAAGAACTCTGTTCAGTCCCTC 1658
Db 41 GCATTATGCTAGGCTGTGCAAGAACTCTGTTCAGTCCCTC 1
RESULT 7
US-09-764-868-522
; Sequence 522, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 522
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (516)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (552)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (558)
; OTHER INFORMATION: n equals a.t,g, or c
US-09-764-868-522

Query Match 19.6%; Score 335.4; DB 9; Length 572;
Best Local Similarity 97.7%; Pred. No. 2.7e-83;
Matches 339; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
Qy 819 CTGTTTCCCTTCGGGGGAAGTCATAGTTAGGAGACCTCACACTTTCCCTCAGAGA 878
Db 165 CTCTGTTTCTAGGCGGGAAGTCATAGTTAGGAGACCTCACACTTTCCCTCAGAGA 224
Qy 879 AGACACACACCAATGACCTCAGCTGGGACCCCACTCCAGGCCCTCAGCCCCCAAC 938
Db 225 AGACACACACCAATGACCTCAGCTGGGACCCCACTCCAGGCCCTCAGCCCCCAAC 284
Qy 939 CTGCCCCAGCCAGCCCTGAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACCCACGGG 998
Db 285 CTGCCCCAGCCAGCCCTGAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACCCACGGG 344
Qy 999 AACTGACTGTGGTCCAGGGAGAGAGCTGGAGTTCTGGACCAAGCGGTGGTGGC 1058
Db 345 AACTGACTGTGGTCCAGGGAGAGAGCTGGAGTTCTGGRCACAGCAAGCGGTGGTGGC 404
Qy 1059 TGGTGAAGAATGAGCGGGAGCGGCTACATTTCGAAGCAACATCTCGGAGCCCTAC 1118
Db 405 TGGTGAAGAATGAGCGGGAGCGGCTACATTTCGAAGCAACATCTCGGAGCCCTAC 464

Qy 1119 AGCCGGGGACCCCTGGGACCCAGGCGCAGTCAACCCTCTCGGGTTCCA 1165
Db 465 AGCCGGGGACCCCTGGGACCCAGGCGCAGTCAACCCTCTCGGGTACTA 511
RESULT 8
US-09-764-868-1345/c
; Sequence 1345, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1345
; LENGTH: 2643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1345
Query Match 18.9%; Score 324; DB 9; Length 2643;
Best Local Similarity 100.0%; Pred. No. 8.3e-80;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1366 GATAAGCCCTTAGCCACGCTTAGACACCTCCAAGAACCCGCGCTGATGCAAGAT 1425
Db 324 GATAAGCCCTTAGCCACGCTTAGACACCTCCAAGAACCCGCGCTGATGCAAGAT 265
Qy 1426 GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTTCTGGATCCAGTTGCAGCA 1485
Db 264 GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTTCTGGATCCAGTTGCAGCA 205
Qy 1486 AACCCACACCCAGCTCACACAGCAAAACAATGGACGCCCGAGGCTGAAGCAAAAC 1545
Db 204 AACCCACACCCAGCTCACACAGCAAAACAATGGACGCCCGAGGCTGAAGCAAAAC 145
Qy 1546 AGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAAACACCTATTTTACCTCTTT 1605
Db 144 AGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAAACACCTATTTTACCTCTTT 85
Qy 1606 CCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCACTCCTCTCTCT 1665
Db 84 CCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCACTCCTCTCTCT 25
Qy 1666 CAATAAAGCATCTTCAAGCTGT 1689
Db 24 CAATAAAGCATCTTCAAGCTGT 1
RESULT 9
US-09-925-299-321
; Sequence 321, Application US/09925299
; Patent No. US2002055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 321
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Homo sapiens


```
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2415
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2415

Query Match          3.5%; Score 59; DB 10; Length 245;
Best Local Similarity 96.0%; Pred. No. 8.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 GCACAGCGACTGAACACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAA---AG 57
    |||||
Db 75 GCACAGCGACTGAACACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAAGCAG 16
    |||||
QY 58 ACCTCGACTTGGAGG 72
    |||||
Db 15 ACCTCGACTTGGAGG 1

RESULT 12
US-09-764-868-99
; Sequence 99, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 1639
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-99

Query Match          3.3%; Score 56.4; DB 9; Length 1639;
Best Local Similarity 54.3%; Pred. No. 1.2e-05;
Matches 114; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 950 GCCTCGAAAATGCAAGTCTTGACGAGTTTGAAGCTAGGAACCCACGCGGAAGTGAAGTGTG 1009
    |||
Db 186 GCAAGTACGCTCAAGATCCTGTATGACTTCACAGCCCGAAATGCCAACGAGCTATCGGTG 245
    |||
QY 1010 GTCCAGGGAGAGAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGAAGAAT 1069
    |||
Db 246 CTCAGGATGAGGTCCTAGAGGTGCTGGAGGCGCGGCGGAGTGGTGAAGCTGCGCAGC 305
    |||
QY 1070 GAGGCGGGAGGCGGCTACATTCCAAAGCAACATCCTCGAGCCCTACAGCGGGGACC 1129
    |||
Db 306 CGCAGGCGGCGGCGGTGACGTGCCCTGCAACATCCTAGGCGGCGGCGGAGGAGC 365
    |||
QY 1130 CTTGGGACCCAGGCGGAGTCAACCTCTCGG 1159
    |||
Db 366 GCGGCGGCGGCTTCGAGGAGCGGCGGTCTAG 395
    |||

RESULT 13
US-09-764-868-512
; Sequence 512, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 512
; LENGTH: 1661
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-512

Query Match          3.3%; Score 56.4; DB 9; Length 1661;
Best Local Similarity 54.3%; Pred. No. 1.2e-05;
Matches 114; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 950 GCCTCGAAAATGCAAGTCTTGACGAGTTTGAAGCTAGGAACCCACGCGGAAGTGAAGTGTG 1009
    |||
Db 175 GCAAGTACGCTCAAGATCCTGTATGACTTCACAGCCCGAAATGCCAACGAGTATCGGTG 234
    |||
QY 1010 GTCCAGGGAGAGAGCTGGAGGTTCTGGACCACAGCAACGCGGTGGTGGTGAAGAAT 1069
    |||
Db 235 CTCAGGATGAGGTCTCTAGAGGTGCTGGAGGAGCGGCGGAGTGGTGAAGTGGCCAGC 294
    |||
QY 1070 GAGGCGGGAGCGGAGCGGCTACATTCCAAAGCAACATCCTCGAGCCCTACAGCGGGGACC 1129
    |||
Db 295 CGCAGCGGCGGAGCGGCGGTACGTGCCCTGCAACATCCTAGGCGGAGCGGCGGAGGAC 354
    |||
QY 1130 CTTGGGACCCAGGCGGAGTCAACCTCTCGG 1159
    |||
Db 355 GCGGCGGCGGCTTCGAGGAGCGGCGGTCTAG 384
    |||

RESULT 14
US-09-833-790-43
; Sequence 43, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-790-43

Query Match          3.2%; Score 55; DB 10; Length 514;
Best Local Similarity 49.2%; Pred. No. 1.6e-05;
Matches 145; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 450 TTGACTGTCTCCAGAAGATCAAGTACAGCTTCAACCTCTGGGAAGGCTGGCCACCTGCG 509
    |||
Db 44 TTGACTGTCTCCAAAAGTTTAAACAGGATTTAAACCTCTGGCCAAACTGAAGTCTCATA 103
    |||
QY 510 TGAAGGAGACAAGTGCCTGAGCTGCTGTACATCCTCTTCAAGTCCCTGAAGTTCATCC 569
    |||
Db 104 TTCAGAAATCCTAGTGTCTGAGATTTGGTTTCACTTTTGTACTCCATTAATATATGGTGG 163
    |||
QY 570 TGGCCAGGTGCCCTGAGGCTGGCTAGCAGCCCAAGTGTATCAACCTCTCCCTCCACCCCTA 629
    |||
Db 164 TGACGCAACAGAGGAGTCTGAACCTAGCCAGTTCAGTACTTAGTCCCTCTATTGAATAAG 223
    |||
QY 630 AAGCTATCAACCTGTCTACAGTCTCTGTCTAAGCCCACTCAGAGTAACTTTGGATGGGT 689
    |||
Db 224 ACACAATTGATTCTTTAAATTTATCTGTCAATGGTGTGAACGGCAGCTGTGGATGTCAT 283
    |||
QY 690 TGGGCCCCAGCCTGGACCACTACCGCGGCGGAGTGGACAGGCGATGAGCCCTGCC 744
    |||
```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 16:13:42 ; Search time 18.5 Seconds
(without alignments)
7667.516 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 3127

Sequence: 1 ggcagagcagctgaagaccca.....aaaaaaaaaataangataaa 1710

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09762021/runat_24022003_153038_7545/app_query.fasta_1.1863
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsf -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021.ecgn_1_1_12_erunat_24022003_153038_7545 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	558.5	17.9	821	1 EPS8_MOUSE	Q08509 mus musculus
2	550	17.6	822	1 EPS8_HUMAN	Q12929 homo sapien
3	217	6.9	1484	1 CA13_MOUSE	P08121 mus musculus
4	200.5	6.4	1486	1 CA13_HUMAN	P02461 homo sapien
5	196.5	6.3	2161	1 SHK1_HUMAN	Q9y566 homo sapien
6	192	6.1	660	1 YHL1_EBV	P03181 epstein-bar
7	192	6.1	2167	1 SHK1_RAT	Q9wv48 rattus norv
8	186.5	6.0	1083	1 T2D3_HUMAN	O00268 homo sapien
9	185	5.9	963	1 YQ36_CAEEL	Q09457 caenorhabdi
10	184.5	5.9	1049	1 SHK1_BOVIN	P04258 bos taurus
11	184.5	5.9	1183	1 DRPL_RAT	P54258 rattus norv
12	184	5.9	2944	1 CA17_HUMAN	Q02388 homo sapien
13	183.5	5.9	2774	1 MAPA_RAT	P34926 rattus norv
14	181	5.8	1453	1 CA11_CHICK	P02457 gallus gall
15	180	5.8	1453	1 CA11_MOUSE	P11087 mus musculus
16	177.5	5.7	547	1 Z213_HUMAN	O14771 homo sapien
17	177	5.7	1729	1 TABP_HUMAN	Q9c0c2 homo sapien
18	174.5	5.6	1464	1 CA11_HUMAN	P02452 homo sapien

19	172.5	5.5	1185	1 DRPL_HUMAN	P54259 homo sapien
20	172	5.5	707	1 SFPO_HUMAN	P23246 homo sapien
21	171.5	5.5	1460	1 CA11_CANFA	Q9x517 canis fami
22	170	5.4	1516	1 CA1H_HUMAN	P39060 homo sapien
23	170	5.4	3530	1 MY15_HUMAN	Q9ukn7 homo sapien
24	168.5	5.4	487	1 EBN2_EBV	P12978 epstein-bar
25	168.5	5.4	671	1 VINE_HUMAN	O60504 homo sapien
26	168.5	5.4	2142	1 BAT2_HUMAN	P48634 homo sapien
27	166	5.3	1418	1 CA12_HUMAN	P02458 homo sapien
28	165	5.3	779	1 CA11_BOVIN	P02453 bos taurus
29	163.5	5.2	1336	1 W146_HUMAN	Q9c0j8 homo sapien
30	163.5	5.2	1459	1 CA12_MOUSE	P28481 mus musculu
31	163	5.2	680	1 CA1A_MOUSE	Q05306 mus musculu
32	162.5	5.2	636	1 CA13_RAT	P13941 rattus norv
33	162.5	5.2	1362	1 CA21_CHICK	P02467 gallus gall
34	162.5	5.2	3149	1 TEGU_EBV	P03186 epstein-bar
35	161.5	5.2	1804	1 CA1B_MOUSE	Q61245 mus musculu
36	161.5	5.2	3511	1 MY15_MOUSE	Q9qz24 mus musculu
37	161	5.1	699	1 VGLG_HSV2H	P13290 herpes simp
38	160.5	5.1	1364	1 CA21_BOVIN	P02465 bos taurus
39	160	5.1	1262	1 CA13_CHICK	P12105 gallus gall
40	158	5.1	867	1 SSPQ_BOVIN	P98167 bos taurus
41	158	5.1	1372	1 CA21_RAT	P02466 rattus norv
42	158	5.1	1838	1 CA15_HUMAN	P20908 homo sapien
43	157.5	5.0	1366	1 CA21_CANFA	O46392 canis fami
44	157	5.0	467	1 CBPA_DICDI	P35085 dictyosteli
45	157	5.0	1736	1 CA2B_HUMAN	P13942 homo sapien

ALIGNMENTS

RESULT 1

EPS8_MOUSE

ID EPS8_MOUSE STANDARD; PRT; 821 AA.

AC Q08509;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Epidermal growth factor receptor kinase substrate EPS8.

GN EPS8.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94008987; PubMed=8404850;

RA Fazioli F., Minichiello L., Matoska V., Castagnino P., Miki T.,

RA Wong W.T., di Fiore P.P.;

RT "Eps8, a substrate for the epidermal growth factor receptor kinase,

RT enhances EGF-dependent mitogenic signals.";

RL EMBO J. 12:3799-3808(1993).

RN [2]

RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 532-591.

RX MEDLINE=97448677; PubMed=9303002;

RA Kishan K.V.R., Scita G., Wong W.T., di Fiore P.P., Newcomer M.E.;

RT "The SH3 domain of Eps8 exists as a novel intertwined dimer.";

CC Nat. Struct. Biol. 4:739-743(1997).

CC -!- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT

CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.

CC -!- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.

CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN. IT IS SPLIT IN TWO PARTS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; L21671; AAA16358.1; -.

DR PDB: IAOJ; 08-JUL-98.
DR MGD; MGI:104684; Eps8.
DR InterPro: IPR000050; PID_domain.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00462; PTB; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00002; SH3; 1.
KW SH3 domain; Phosphorylation; 3D-structure.
FT DOMAIN 69 129
FT DOMAIN 210 213 PH (FIRST PART).
FT DOMAIN 322 325 POLY-PRO.
FT DOMAIN 381 414 PH (SECOND PART).
FT DOMAIN 421 440 PRO-RICH.
FT DOMAIN 532 591 SH3.
FT DOMAIN 620 650 PRO-RICH.
FT DOMAIN 658 663 POLY-SER.
SQ SEQUENCE 821 AA; 91738 MW; 6B9EB95DD22D910C CRC64;

Alignment Scores:
Pred. No.: 5, 55e-25 Length: 821
Score: 558.50 Matches: 158
Percent Similarity: 38.82% Conservative: 78
Best local Similarity: 25.99% Mismatches: 171
Query Match: 17.86% Indels: 201
DB: 1 Gaps: 17

US-09-762-021a-1 (1-1710) x EPS8_MOUSE (1-821)
QY 101 GGGCTGTATGAAAGCGCGCTCCCTATGGAGCAGCGCATCTGGAGCGGGGATC 160
DB 190 GlyLysGlnLysArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209
QY 161 CCTCCAGACAGCCGCCACAGGAGCCCTAGACAGCCCTCCACCATCCCAAGGCC 220
DB 210 ProProPro-----ProArgAlaProAlaPro 218
QY 221 CTGCGACGCCACACC-----AGTGCC 241
DB 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla 238
QY 242 CGAGAACCAAGTCTTACTCTGCTCTCCCAAGCGGCTCTTCCCGGAGGAC--- 298
DB 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
QY 299 CCAGAG-----AGGACGAGGAGTCTGCAACCATCTCTCCTAAGG 337
DB 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
QY 338 GACATTGAGCTGTCATGGAAAGCTGGAGAGGCCAG-----GCA 379
DB 279 AspileGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
QY 380 AGACACAGCAGGAGAAATTTGGGAAAAAACAAGGACCGAGGAGGTCTCACC--- 436
DB 299 LysArgLysLysSerLysLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
QY 437 -----CAGGCACAGTACATTGACTGCTCCCAAGATCAAGTACAGC 478
DB 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
QY 479 TTTCAACCTCTGGGAGGCTGCCACCTGGCTGAAGGAGACAAAGTGCCTCAGCTCCTA 538
DB 339 PheAsnLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
QY 539 CACATCTCTTCAAGTCCCTGAATTCATCTCGGCCAGGTCGCCCTGAGGCTGGCTAGCA 598
DB 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
QY 599 GCCCAAGTATCTACCCCTCTCACCCTTAAGCTATCAACCTGCTACAGTCTCTGTCTA 658
DB 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398

QY 659 AGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGACCACCTAGCGGGCC 718
DB 399 ThrAlaGluAlaArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
QY 719 GACTGGACAGCGATGAGCCCTG---CCCTACCAACCCACATTTCTCAGATGACTGGCAA 775
DB 419 GluTrpProLysGluGlnPheIleProProTyrValProArgPheArgAsnGlyTrpGlu 438
QY 776 CTTCCTCA-----GAGCCCTCCAGCCCAAGCACCCTTAGGTAGTACCAGAC 817
DB 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
QY 818 CCGTGTTCCTCCCTCGGGGGAAGTCAT---AGGTTTAGGAGCAGCTCACACTTCTCCTCAG 874
DB 457 AlaGluSerValAlaAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
QY 875 GAGAAGACACAAACCATGACCTCAGCTCGGGAC----- 910
DB 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerMet 496
QY 911 -----CCC 913
DB 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
QY 914 AAC-----TCCAGGCCCTCCAGGCCCAACCTGCCAGCCAGCCCTGAAATGCAA 964
DB 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
QY 965 GTCTGTGTACAGTTGAAGCTAGGAAACCCCGGAACCTAGTGTGTCCAGGGAGAGAG 1024
DB 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal 556
QY 1025 CTGAGGTTCTCGACACAGCAAGCGTGTGGTGTGAACAATCAGCCGGGAGGAGC 1084
DB 557 LeuGluIleLeuAspAspArgArgGlnTrpLysValArgAlaSerGlyAspSer 576
QY 1085 GGCTACATTTCCAAGCACATCTCGAG----- 1111
DB 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
QY 1111 ----- 1111
DB 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
QY 1112 -----CCCCCTA 1117
DB 617 AlaAspThrProSerAlaProSerProProThrProAlaProValProValProLeu 636
QY 1118 CAGCGGGGAGCCCTCGG----- 1135
DB 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
QY 1135 ----- 1135
DB 657 AsnSerSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
QY 1135 ----- 1135
DB 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
QY 1136 ---ACCCAGGCGAGCTCACCTCTCGG-----GTTCCA 1165
DB 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
QY 1166 ATGCTTCGACTTAGC-----TCGAGGCGCTGAAGGTCACAGACTGCTGCGAGCAGAG 1219
DB 717 ValIleAsnIleThrTyrAspSerProGluGluValLysThrTrpLeuGlnSerLys 736
QY 1220 AACTTCTCCACTGCCCGTGGAGACACTTGGTCCCTCAGCGGGGAGCCAGCTACTTCGC 1279
DB 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756

```
Qy 1280 ATAGACCTGGGGAGCTACAGATGCTATGTGTCACAGAGGCCCCACAGATCTCTCCCGG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValCysPro---GluGlyAlaArgValPheAsnGln 775
Qy 1340 CTGGAGGCTGTCAGAGGATGCTG 1363
Db 776 IieThrValGlnLysAlaLeu 783

RESULT 2
EPS8_HUMAN STANDARD; PRT; 822 AA.
AC Q12929;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor kinase substrate EPS8.
GN EPS8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94366758; PubMed=8084614;
RA Wong W.T., Carlomagno F., Druck T., Barletta C., Croce C.M.,
RA Huebner K., Kraus M.H., di Fiore P.P.;
RT "Evolutionary conservation of the EPS8 gene and its mapping to human
RT chromosome 12q23-q24."
RL Oncogene 9:3057-3061(1994).
CC -1- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT
CC MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES ANALYZED, INCLUDING
CC HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND
CC PANCREAS. EXPRESSED IN ALL EPITHELIAL AND FIBROBLASTIC LINES
CC EXAMINED AND IN SOME, BUT NOT ALL, HEMATOPOIETIC CELLS.
CC -1- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN. IT IS SPLIT IN TWO PARTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12535; AAA62280.1; -.
DR HSP; Q08509; IAOJ.
DR Genew; HGNC:3420; EPS8.
DR MIM; 600206; -.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain; Phosphorylation.
FT DOMAIN 69 129 PH (FIRST PART).
FT DOMAIN 210 213 POLY-PRO.
FT DOMAIN 322 325 POLY-PRO.
FT DOMAIN 381 414 PH (SECOND PART).
FT DOMAIN 421 440 PRO-RICH.
FT DOMAIN 532 591 SH3.
FT DOMAIN 615 651 PRO-RICH.
FT DOMAIN 659 664 POLY-SER.
SQ SEQUENCE 822 AA; 91881 MW; AC5EB1D28B784B3B CRC64;

Alignment Scores:
Pred. No.: 1.7e-24 Length: 822
Score: 550.00 Matches: 158
Percent Similarity: 38.71% Conservative: 77
```

```
Best Local Similarity: 26.03% Mismatches: 174
Query Match: 17.59% Indels: 198
DB: 1 Gaps: 15

US-09-762-021a-1 (1-1710) x EPS8_HUMAN (1-822)
Qy 101 GGGCTCTATGAAAGCGCGCTCCCTATATGAGCAGCAGCGCTATCTGGAGCCGGGATC 160
Db 190 GlyLysGlnLysArgProAspAlaLeuArgMetIleSerAsnAlaAspProSerIle 209
Qy 161 CCTCCAGAACAGCCCCAGCAGAGGAGCCCTAGAGCAGCAGCTCCACCACATCCCAAGGCC 220
Db 210 ProProPro-----ProArgAlaProAlaPro 218
Qy 221 CTGCCAGCCGACACAC-----AGTGCC 241
Db 219 AlaProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla 238
Qy 242 CGAAGAACCAAGTGCCTTTACTCTCCCTCCCAAGGGCGTCTCTCCCGGAGGAC--- 298
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
Qy 299 CCAGAG-----AGGACGAGGAAGTGTGAACCATGTCTCTAAGG 337
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
Qy 338 GACATTGAGTGTTCATGGGAAGCTGGAGAAGGCCAG-----GCA 379
Db 279 AspIleGluPheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
Qy 380 AAGACCAAGCAAGAAAGAAATTTGGGAAAAAACAAGGACCAAGGAGGTCTCACC--- 436
Db 299 LysArgLysLysAsnLysLysGlyLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
Qy 437 -----CAGGCACACTACATTGCTCTCCAGAGATCAAGTACAGC 478
Db 319 ArgAlaLysProProProAspPheLeuAspCysPheGlnLysPheLysHisGly 338
Qy 479 TTCACCTCTCGGAAGCTGGCCACCTGGCTGAAGGAGACAAGTCCCTGACCTCGTA 538
Db 339 PheAsnLeuLeuAlaLysLysLysSerHisIleGlnAsnProSerAlaAlaAspLeuVal 358
Qy 539 CACATCTCTCTCAAGTCCCTGAACTTCATCTCGCCAGGTGCCCTGAGGTGGCTAGCA 598
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
Qy 599 GCCCAAGTATCTCACCCCTCTCCCTCCCTAAAGCTATCAACCTGCTACAGTCTCTGCTA 658
Db 379 SerSerValLeuSerProLeuLeuAsnLysAspThrIleAspPheLeuAsnTyrThrVal 398
Qy 659 AGCCCACTGAGAGTAACCTTTGGATGGGTGGGCCAGCTGGACCACTAGCGGGGCC 718
Db 399 AsnGlyAspGluArgGlnLeuTrpMetSerLeuGlyGlyThrTrpMetLysAlaArgAla 418
Qy 719 GACTGGACAGCGGATGAGCCCGCTG---CCCTACCAACCCACATTTCTCAGACTGGCAA 775
Db 419 GluTrpProLysGluGlnPheIleProProTyrValProArgPheArgAsnGlyTrpGlu 438
Qy 776 CTTCAGAGCCCTCCAGCCCAAGCACCTTAGGATACAGGACCCCTGTTCCCTT----- 829
Db 439 ProProMetLeuAsnPheMetGlyAlaThrMetGluGlnAspLeuTyrGlnLeuAlaGlu 458
Qy 830 -----CGCGGGGAAGTCAATAGTTAGGAGCAGCACCTCA 862
Db 459 SerValAlaAsnValAlaGluHisGlnArgLysGlnGluIleLysArgLeuSerThrGlu 478
Qy 863 CACTTTCTCAGGAGAACACACACCATGACCTCAGCCCTGGGGACCCCAAC----- 916
Db 479 HisSerValSerGluTyrHisProAlaAspGlyTyrAlaPheSerSerAsnIleTyr 498
Qy 917 -----TCCAGGCCCTCCAGC----- 931
Db 499 ThrArgGlySerHisLeuAspGlnGlyAlaAlaValAlaPheLysProThrSerAsn 518
```

```
QY 932 -----CCCAACCTGCCAGCCAGCCCTGAAATGCAAGTC 967
DB 519 ArgHisIleAspArgAsnTyrGluProLeuLysThrGlnProLysTyrAlaLysSer 538
QY 968 TTGTACGAGTTTGAAGCTAGGAAACCCAGGGAAGTCTGTGCTCCAGGGAGAGAGCTG 1027
DB 539 LysTyrAspPheValAlaArgAsnSerGluLeuSerValLeuLysAspIleLeu 558
QY 1028 GAGGTCTCGACACAGCAAGCGGTGGTGGTGAAGATGAGCGGAGCGGAGCGCGC 1087
DB 559 GluIleLeuAspArgLysGlnThrPrpLysValArgAsnAlaSerGlyAspSerGly 578
QY 1088 TACATTCGAAGCAACACTCTGGAG----- 1111
DB 579 PheValProAsnAsnIleLeuAspIleValArgProProGluSerGlyLeuGlyArgAla 598
QY 1111 ----- 1111
DB 599 AspProProTyrThrHisThrIleGlnLysGlnArgMetGluTyrGlyProArgProAla 618
QY 1112 -----CCCTACAG 1120
DB 619 AspThrProAlaProSerProProThrProAlaProValProValProLeuPro 638
QY 1121 CCGGGAGCCCTCTGGG----- 1135
DB 639 ProSerThrProAlaProValProValSerLysValProAlaAsnIleThrArgGlnAsn 658
QY 1135 ----- 1135
DB 659 SerSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgHisLysGlnLeu 678
QY 1135 ----- 1135
DB 679 ProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuIleHisArgLeu 698
QY 1136 ACCAGGGCCAGTCACTCCCTCTCGG-----GTTCCAAATG 1168
DB 699 ThrIleGlyArgSerAlaAlaGlnLysLysPheHisValProArgGlnAsnValProVal 718
QY 1169 CTTCACTTAGC-----TCGAGGCTGAAGAGTCAAGAGTCTGCTGCAGGCAGAGAAC 1222
DB 719 IleAsnIleThrTyrAspSerThrProGluAspValLysThrTrpLeuGlnSerLysGly 738
QY 1223 TTCTCCACTGCCAGCTGAGGACACTTGGTCCCTGACGGGAGCAGCTACTTCGCATA 1282
DB 739 PheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSerLeu 758
QY 1283 AGACCTGGGAGCTACAGTGTATGTCCACAGAGGCGCCACGAATCCTGTCCCGGCTG 1342
DB 759 AsnLysAspGluLeuArgThrValCysPro---GluGlyAlaArgValTyrSerGlnIle 777
QY 1343 GAGGCTGTCAAGAGATCGTG 1363
DB 778 ThrValGlnLysAlaAlaLeu 784
RESULT 3
CAL3_MOUSE
ID CAL3_MOUSE STANDARD; PRT: 1464 AA.
AC P08121; Q61429; Q9CRN7;
DT 01-JUL-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
```

```
RX MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RL complete DNA sequence.";
RN Gene 147:161-168(1994).
[2]
RX SEQUENCE OF 1-488 FROM N.A.
RP MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RL alpha-1 type-III collagen chain.";
RN Gene 61:225-230(1987).
[3]
RX SEQUENCE OF 1-28 FROM N.A.
RP MEDLINE=85131189; PubMed=3972847;
RA Liau G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RL (III) collagen gene.";
RN J. Biol. Chem. 260:3773-3777(1985).
[4]
RX SEQUENCE OF 810-1464 FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Embryonic head;
RL MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikolaio I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
[5]
RX SEQUENCE OF 1442-1464 FROM N.A.
RP STRAIN=C57BL/6;
RL MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RL collagen mRNAs.";
RN Biochim. Biophys. Acta 1089:241-243(1991).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; X52046; CAA36279.1; -
DR EMBL; M18933; AAA37338.1; -
DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR EMBL; AK019448; BAB31724.1; -
```


DR EMBL; X57983; CAA41048.1; -.
 DR PIR; A22287; A22287.
 DR PIR; A27353; A27353.
 DR PIR; S16373; S16373.
 DR MGI; 88453; Col3a1.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF01410; COLFI; 1.
 DR ProDom; PD000007; Collagen; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90 VWF.
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
 FT DOMAIN 1196 1464 COLLAGEN ALPHA 1(III) CHAIN.
 FT CARBOHYD 262 262 O-LINKED (GAL. . .) (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 976 976 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
 FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Alignment Scores:
 Pred. No.: 1,75e-05 Length: 1464
 Score: 217.00 Matches: 156
 Percent Similarity: 31.04% Conservative: 38
 Best Local Similarity: 24.36% Mismatches: 218
 Query Match: 6.94% Indels: 213
 DB: 1 Gaps: 34

US-09-762-021A-1 (1-1710) x CA13_MOUSE (1-1464)

Qy 10 ACTGAAGACCCCTGCAGAA---GGCTCTGGAGGAAGAGCTGGAGCAAAACACCTCGACT 66
 Db 645 ThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluVal 664
 Qy 67 TGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGCTCTATGGAAGGCCCTCCC 126
 Db 665 GlyAlaProGlyAlaProGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro 684
 Qy 127 TATGGAGCA-----GGCAGCGTATCTGGAGCGGGGATCCCTCCAGAACACGCC 174
 Db 685 ProGlyThrAlaGlyLeuProGlyAlaArgGlyGlyAlaGlyProGly----- 701
 Qy 175 CCACAGAGGACCCCTAGA-----GCACAGCGCTCCACCATCCCAAGGCCCTGCCACG 228
 Db 702 ---ProGlyGlyGlyLysGlyProAlaGlyProProGlyProGlyAlaSerGlySer 720
 Qy 229 CCACAGCAGTCCCGAGAACCAAGTGCCTTTACTCTGCCTCTCCAGCGCTCTCTC 288
 Db 721 Pro-----GlyLeuGlnGly 725
 Qy 289 CCCCAGGACCCAGAGAGGAGGAGAAAGTGTGAACCATGCTCTAAGGGACATTGAGCT 348
 Db 726 MetProGlyGluArgGlyGlyProGlySer-----ProGlyProLysGlyGlyLysGly 743
 Qy 349 GTTCATGGGAAGCTGGA---GAAGGC---CCAGCAAGACACGACGAGAAAGAAAT 402
 Db 744 GluProGlyGlyAlaGlyAlaAspGlyValProGlyLysAsp----- 757

Qy 403 TGGGAAAAAACAAGGACCA---GGGAGGTCTCACCAGGACACAGTACATTGACTGCTT 459
 Db 758 -----GlyProArgGlyProAlaGlyPro----- 765
 Qy 460 CCAGAAGATCAAGTACAGCTTCAACCTCCTGGGAGGCTGGCCACCTGGCTGAAGGAGAC 519
 Db 766 -----IleGlyProProGlyProAlaGlyGln-----ProGlyAsp 777
 Qy 520 AAGTGCCTCAGCTGCTGCTACACATCTCTCAAGTCCCTGAACCTCATCTCTGGCCAGGTG 579
 Db 778 LysGlyGluGlyGlySer-----ProGlyLeu---ProGlyLeuAla 790
 Qy 580 ---CCCTGAGGCTGGCTAGCAGCCCAAGTATGATCTCAACCTCCCTCACCCTAAAGCTAT 636
 Db 791 GlyProArgGlyGlyProGlyGluArgGlyGluHisGlyProProGlyProAlaGlyPhe 810
 Qy 637 CAACCTGCTACAGTCTGTCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTG----- 691
 Db 811 ProGlyAlaProGlyGlnAsn-GlyGluProGlyAlaLysGlyGluArgGlyAlaProGly 830
 Qy 692 -----GGCCAGCTGGACCACTAGCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 735
 Db 830 yGluLysGlyGluGlyGlyPro-ProGlyPro---AlaGlyProThrGlySerSerGlyP 849
 Qy 736 GCCCTGCCTACCAACCCACATCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCA 795
 Db 849 roAlaGlyProProGlyProGlnGly---ValLysGlyGluArgGlySerPro----- 865
 Qy 796 AGCACCTTAGGATACAGGACCCCTGTTTCCCTTCGGCGGGGAGTATAGTTAGGAG 855
 Db 866 -----GlyG 867
 Qy 856 CACCTCACACTTCTCTCAGGAGAAGACACAAACCATCACCCTCAGCCTGGGGACCCA- 914
 Db 867 lyProGlyThr-----AlaGlyPheProG 875
 Qy 915 -----ACTCCAGGCGCTCCAGCCCAACCTCCAGCCAGCCAGCCGAAATGCA 963
 Db 875 lyGlyArgGlyLeuProGlyProProGlyAsnAsn-----GlyAsnPro----- 889
 Qy 964 AGTCTGTACAGTTTGAAGCTAGGAACCCAGGGAACCTGCTGTGTCAGGAGAGAA 1023
 Db 890 -----Gly-ProProGlyProSerGlyAlaProGlyLysAsp 901
 Qy 1024 -----GCTGGAGTTCTGGA---CCACAGCAAGCGGTGGTGGCTGCTGAA 1065
 Db 902 GlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyLeuAlaGlyPro 921
 Qy 1066 GAATGAGCGGGAGCGGCTACATTCCAAGCAACATCTCTGGAGCCCTTACAGCCGGG 1125
 Db 922 Lys---GlyAspAlaGlyGlnProGlyGlyLysGlyProGlyAla-----GlnGly 938
 Qy 1126 GACCCCTGGGACCCAGGCGGAGTCAACCTCTCGGTTTCCAATGCTTCGACTTAGCTC- 1182
 Db 939 ProProGlySerProGlyProLeuGlyLeuAlaGlyLeuThrGlyAlaArgGlyLeuAla 958
 Qy 1183 -----GAGGCTGAAGAGGTACAGACTGGCT 1209
 Db 959 GlyProProGlyMetProGlyProArgGlySerProGlyProGlnGlyLeuLysGlyGlu 978
 Qy 1210 GCAGCAGAGAATCTTCCACTGCCAGCGGTGAGGACACTTGGTCCCT- 1257
 Db 979 SerGlyLysProGlyAlaSerGlyHisAsnGlyGluArgGlyProProGlyProGlnGly 998
 Qy 1258 -----GAC 1260
 Db 999 LeuProGlyGlnProGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAsp 1018
 Qy 1261 GGGAGACCGAGCTACTTCCATTAAGACCTGGGAGGCTACAGATGCTATGTCACAGCA- 1317
 Db 1019 GlyGlnProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluAsnGlySer 1038

QY 1318 -----GGCCCCAGAAATCCTGTCTCCCGCTGGAGCC 1347
 Db 1039 ProGlyAlaProGlyAlaProGlyHisProGlyProGlyProValGlyProSerGly 1058
 QY 1348 TGTCAAGAGTCTGGGGATAG-----CCCTTAGGCACACAGCTTAGACACC 1395
 Db 1059 LysSerGlyAspArgGlyGluThrGlyProAlaGlyProSerGlyAlaProGlyProAla 1078
 QY 1396 TCCAAGAACAGCCCGCTGATGCAAGATGCGAGATCTGTATACCATTTAGAGCCCGAG 1455
 Db 1079 GlyAlaArgGlyAlaProGlyProGlnGlyProArgGlyAspLysGlyGluThrGlyGlu 1098
 QY 1456 AATTCCTCTCTGGATCCAG-----TTTGCAGCAAAACCCACACCCCGAGTCTC 1503
 Db 1099 ArgGlySerAsnGlyIleLysGlyHisArgGlyPheProGlyAsnProGlyPro----- 1116
 QY 1504 ACACAGCAAAACATGGACGCGCCAGAGCTGAAGCAACAGTGTCTCTCTGCGTGT 1563
 Db 1117 -----ProGlySerProGlyAlaAlaGlyHisGlnGlyAla 1128
 QY 1564 GTTGAGCCTCCCGAGTAACACCATTTATTTTACCTCTTTCCCAAACTGGAGCATTT 1623
 Db 1129 IleGlySer-----ProGlySer-----ProGlyProAla 1135
 QY 1624 ATGCCCTAGG 1632
 Db 1136 GlyProArg 1138
 RESULT 4
 CA13_HUMAN
 ID CA13_HUMAN STANDARD; PRT; 1466 AA.
 AC P02461; Q15112;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Prockop D.J.;
 RA "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 [2]
 RN SEQUENCE OF 149-1225 FROM N.A.
 RP MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RA "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 [3]
 RN SEQUENCE OF 168-398.
 RP MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RA "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 [4]
 RN REVISIONS.
 RP Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 [5]
 RN SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;

RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBR peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 [6]
 RN SEQUENCE OF 728-964.
 RP MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RA "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 [7]
 RN SEQUENCE OF 950-1466 FROM N.A.
 RP MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Dalgleish R.;
 RA "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 [8]
 RN REVISION TO 1184.
 RP MEDLINE=89098346; PubMed=3211760;
 RA Molyneux K., Dalgleish R.;
 RA "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 [9]
 RN SEQUENCE OF 1065-1466 FROM N.A.
 RP MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenbloom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 [10]
 RN SEQUENCE OF 965-1200.
 RP MEDLINE=81208139; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB9 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 [11]
 RN SEQUENCE OF 1176-1466 FROM N.A.
 RP MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RA "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 [12]
 RN SEQUENCE OF 1161-1200 FROM N.A.
 RP MEDLINE=86187804; PubMed=3754462;
 RA Miskulin M., Dalgleish R., Klueve-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 [13]
 RN SEQUENCE OF 1-170 FROM N.A.
 RP TISSUE=Placenta;
 RX MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human prepro alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 [14]
 RN SEQUENCE OF 1-176 FROM N.A.
 RP MEDLINE=8938752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 [15]
 RN REVIEW ON VARIANTS.
 RP MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;


```
QY 742 GCCTTACCACCCACATCTTCAGATGACTGGCAACTTCAGAGCCCT----- 788
DB 852 lyProGlyProGlnGly---ValLysGlyGluArgGlySerProGlyGlyProGlyA 871
QY 789 -----
DB 871 laAlaGlyPheProGlyAlaArgGlyLeuProGlyPro-ArgGlySerAsnGlyAsnPro 890
QY 814 GGACCTGTTTCCTTCGGCGGGGAAGTCATAGTTAGGAGCACCTCACACTTCCTCA 873
DB 891 GlyProGlyProSerGly-----SerPro 899
QY 874 GGAGAAGCACACCAACCATGACCTCA-----GCC 903
DB 900 GlyLysAspGlyProProGlyProAlaGlyAsnThrGlyAlaProGlySerProGlyVal 919
QY 904 TGGGGACCCCACTCCAGGCCCTCCAGCCC-----CAAGCC 939
DB 920 SerGlyProLysGlyAspAlaGlyGlnProGlyGluLysGlySerProGlyAlaGlnGly 939
QY 940 TGCCACGACGACCTCAAAATGCAAGTCTGTACGAGTTTGAAGCTAGGACCCAGGGA 999
DB 940 ProProGlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAla 959
QY 1000 ACTGACTGTGTCCAGGGAGAGAGAGCTGGAGGTTCTGGACACGACGAGCGGTGGTGC 1059
DB 960 GlyProGlyMetProGlyProArgGlySerProGlyProGlnGly-----ValLys 977
QY 1060 GTTGAGATGAGGC-----GGGACGGAGCGGCTACATTCGAAGCAACATCTGTGAGCC 1113
DB 978 GlyGluSer---GlyLysProGlyAlaAsnGlyLeuSerGlyGluArgGlyProProGly 996
QY 1114 CCF-----ACAGCGGGGAGCCCC-----TGGGACCA 1140
DB 997 ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGluProGlyArgAspGlyAsnPro 1016
QY 1141 GGCCCACTCACCCTCGGGTTCCTCAATGCTTCGACTTAGCTCGAGCCCTGAAGAGCTAC 1200
DB 1017 GlySerAspGlyLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGlu 1036
QY 1201 AGACTGGCTGCAGGACAGAACTTCTCACTGCCAGGTGAGGACACTTGGGTCCCTGAC 1260
DB 1037 -----Asn 1037
QY 1261 GGGAGCCAGCTACTTTCGATAGACCTGGGAGCTACAGATGCTATGTCACAGGAGCC 1320
DB 1038 GlySerPro-----GlyAlaProGlyAlaProGlyHisProGly 1050
QY 1321 CCACGAATCTGTCCTCCGCTGGAGCTGTCAAGAGGATGCTGGGGATAAG----- 1371
DB 1051 ProProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluSerGlyProAla 1070
QY 1372 ---CCCTTAGGCACAGCTTAGACACTTCAGAACCCAGCCCGCTGTATGCAAGATGCC 1428
DB 1071 GlyProAlaGlyAlaProGlyProAlaGlySerArgGlyAlaProGlyProGlnGlyPro 1090
QY 1429 AGATCTGATACCATTAGAGCCCGAGAAATTCCTCTCTGATGCCAG----- 1476
DB 1091 ArgGlyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGly 1110
QY 1477 TTTCGACCAACCCC-----ACACCCCACTCACACAGCAACCAACAAATGGACA 1524
DB 1111 PheProGlyAsnProGlyAlaProGlySerPro----- 1121
QY 1525 GGGCCAGAGCTCAACCAACACTGTC-----CCTTCGTGCTGT 1563
DB 1122 GlyProAlaGlyGlnGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyPro 1141
QY 1564 GTTGGAGCCTCC 1575
DB 1142 ValGlyProSer 1145
```

RESULT 5

```
SHK1_HUMAN
ID AC Q9V566; Q9NYW3; STANDARD; PRT; 2161 AA.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1)
DE (Somatostatin receptor interacting protein) (SSTR interacting protein)
DE (SSRIP).
GN SHANK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND INTERACTION WITH SSTR2.
RC TISSUE=Fetal brain, Hippocampus, and Thalamus;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
RL multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).
RN [2]
RP REVIEW.
RX PubMed=10806096;
RA Sheng M., Kim E.;
RL "The Shank family of scaffold proteins.";
CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
CC density (PSD) of excitatory synapses that interconnects receptors
CC of the postsynaptic membrane including NMDA-type and metabotropic
CC glutamate receptors via complexes with GKAP/PSD-95 and Homer,
CC respectively, and the actin-based cytoskeleton. May play a role in
CC the structural and functional organization of the dendritic spine
CC and synaptic junction.
CC -!- SUBUNIT: May homodimerize via its SAM domain (By similarity).
CC Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with
CC SPAN1, Homer-1 and DLGAP1/GKAP isoforms 1 and 2 (By similarity).
CC Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By
CC similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
CC neuronal cells (By similarity).
CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/a (shown here), 2/b and 3; are
CC produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed in brain particularly in the
CC amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
CC seems to be expressed ubiquitously.
CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF163302; AAD45121.1; -.
CC DR EMBL: AF226728; AAF35887.1; -.
CC DR HSP: P06241; 1SHF.
CC DR MIM: 604999; -.
CC DR InterPro: IPR002110; ANK.
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR001660; SAM.
CC DR InterPro: IPR001452; SH3.
CC DR Pfam: PF00023; ank; 7.
CC DR Pfam: PF00595; PDZ; 1.
CC DR Pfam: PF00536; SAM; 1.
```

DR Pfam: PF00018; SH3; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00248; ANK; 3.
 DR SMART: SM00228; PDZ; 1.
 DR SMART: SM00454; SAM; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50088; ANK_REPEAT; 3.
 DR PROSITE: PS50297; ANK_REPEAT; 1.
 DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS50105; SAM_DOMAIN; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW ANK repeat; SH3 domain; Repeat; Alternative splicing.
 FT REPEAT 212 245 ANK 1.
 FT REPEAT 246 278 ANK 2.
 FT REPEAT 279 312 ANK 3.
 FT REPEAT 313 345 ANK 4.
 FT REPEAT 346 378 ANK 5.
 FT REPEAT 379 395 ANK 6.
 FT DOMAIN 554 613 SH3.
 FT DOMAIN 663 757 PDZ.
 FT DOMAIN 2098 2161 SAM.
 FT DOMAIN 1002 1007 POLY-HIS.
 FT DOMAIN 1014 1019 POLY-HIS.
 FT DOMAIN 1189 1195 POLY-GLY.
 FT DOMAIN 1709 1717 POLY-GLY.
 FT DOMAIN 1844 1854 POLY-PRO.
 FT DOMAIN 1896 1902 POLY-GLY.
 FT DOMAIN 1970 1979 POLY-SER.
 FT VARSPPLIC 1 613 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 614 654 RSQESKQESDRKAKLFRHYTVGSDSFSDAPSLMDGIGPG
 FT -> MOLMALEQRFSGPLPGGGOPCLLMSSPLPPPHFSC
 FT LPA (IN ISOFORM 2).
 FT VARSPPLIC 646 654 MISSING (IN ISOFORM 3).
 FT SEQUENCE 2161 AA; 225019 MW; 5FEFC969CBE98701 CRC64;

Alignment Scores:

Pred. No.: 0.00026 Length: 2161
 Score: 196.50 Matches: 174
 Percent Similarity: 30.70% Conservative: 51
 Best Local Similarity: 23.74% Mismatches: 265
 Query Match: 6.28% Indels: 243
 DB: 1 Gaps: 35

US-09-762-021A-1 (1-1710) x SHK1_HUMAN (1-2161)

QY 25 GCAGAGGCTCTGCAGAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCAGCCAGG 84
 DB 1319 AlaTyrGlyGlyGlyGlySerSerAlaAlaPheThrSerPheLeuProProArgProLeu 1338
 QY 85 CCAGGACAGATGAGGGGGCTGTATGTGAAAGGCGCTCCCTATGAGCAGGCG----- 138
 DB 1339 ValHisProLeuThrGlyLysAlaLeuAlaUspproAlaSerProLeuGlyLeuAlaLeuAla 1358
 QY 139 -----ACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCC 174
 DB 1359 AlaArgGluArgAlaLeuLysGluSerSerGlyGlyGlyAlaProGlnProProPro 1378
 QY 175 CCACCAGAGACCTAGAGCAGACGCTCCACACCTCCCAAGGCCCTGCCAGCCACAC 234
 DB 1379 ArgProProSerProArg-TyrGluAlaProPro-----ProThrProHisHisH1 1395
 QY 235 CAGTGCCCGAGAACCAAGTCCTTTACTGCTCTCCCTCCCAAGGGG---TCCTCTCCCC 291
 DB 1395 sSer-----ProHisAlaHisHisGluProValLeuArgLeuTrpGlyAlaSerPr 1412
 QY 292 CGAGGACCCAGAGAGGACGAGGAAGTGCCTGAACCATGTCTTAAGGACATTTAGCTGTT 351
 DB 1412 oProAspProAlaArgArgGluLeuGlyTyrArgAlaGlyLeu-GlySerGlnGluLys 1432
 QY 352 CATGGGAAAGCTGGAGAGGCCCGAGCAAGACAGCAGGAGAAATTTGGGAAAAA 411
 DB 1432 er-----LeuProAlaSerProProAlaAlaArgArgSerLeu-----L 1445

QY 412 AAACAAGGACGAGGAGGTCTCACCCAGGCACAGTACATTGCTTCCAGAGATCAA 471
 DB 1445 euHisArgLeuProProThrAlaProGlyValGlyProLeuLeuGlnLeuGlyThrG 1465
 QY 472 GTACA-----GCTTCAACCTCTGGGAGGCTGCCACCTGGGTGAA 513
 DB 1465 luProProAlaProHisProGlyValSerLysProTrpArgSerAlaAlaProGluGluP 1485
 QY 514 GGAGACAAGTGCCCTG-----AGCTCGTACA 540
 DB 1485 roGluArgLeuProLeuHisValArgPheLeuGluAsnCysGlnProArgAlaProVal 1505
 QY 541 CATCCTCTTCAAGTCCCTGAACCTTCATCTGGCAGGTGCCCTGAGCTGCCCTAGCAGC 600
 DB 1505 hrSerGlyArgGlyProProSerGluAspGlyProGlyVal-----Pro-ProProSer 1522
 QY 601 CCAAGTGATCTACCCCTCTCTACCCCTAAAGCTATCAA----- 639
 DB 1523 ProArgArgSerValProProSerProThrSerProArgAlaSerGluGluAsnGlyLeu 1542
 QY 640 CCTGCTACAGTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGTGGGCCCGCAGC 699
 DB 1543 ProLeuLeuValLeuProPro-----AlaProSer 1553
 QY 700 CTGGACCACTAGCCGGCGCAGCTGGACGCGATGAGCCCTGCCCTTACCAACCCACATT 759
 DB 1554 ValAsp-ValGluAspGlyGluPheLeuPheValGluProLeuProProLeuGluPh 1573
 QY 760 CTCAGATGACTGCAACTTCCAG----- 784
 DB 1573 eSerAsnSerPheGluLysProGluSerProLeuThrProGlyProProHisProLeuPr 1593
 QY 785 -----CCCTCCAGCAACGACCTTAGATACCAAGACCTGTTCCCTCGCGCGGG 837
 DB 1593 oAspThrProAlaProAlaThrProLeuProProValProProAlaValAlaAla 1613
 QY 838 AAGTCATAGTTAGGGAGCAC-----TCACACTTTCCTCAGGAG----- 877
 DB 1613 aProProThrLeuAspSerThrAlaSerSerLeuThrSerTyrAspSerGluValAlaTh 1633
 QY 878 -AAGACACACACCATGACCTCAGCTGGGGACCCCACTCA-----GGCCCTCCAGCCC 933
 DB 1633 rLeuThrGlnGlyAlaSerAlaAlaProGlyAspProHis-ProProGlyProProAlaP 1653
 QY 934 CAACCTGCCAGCC-----AGCCCTGAAATGCAAGCTCTGTGACA 975
 DB 1653 ro-AlaAlaProAlaProAlaAlaProGlnProGlyProAsp----- 1666
 QY 976 GTTTGAAGCTAGGAACCCAGGAACTGACTGT----- 1008
 DB 1667 -----ProProGlyThrAspSerGlyIleGluGluValaspSerArgSer 1682
 QY 1008 ----- 1008
 DB 1683 SerSerAspHisProLeuGluThrIleSerSerAlaSerThrLeuSerSerLeuSerAla 1702
 QY 1009 ---GGTCCAGGAGAGAGCTGGAGTTCTGGACCACAGACGGGTGGTGGTGAA 1065
 DB 1703 GluGlyGlySerAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1722
 QY 1066 GAATGAGCGGGAGC-----GAGCGGTACATTCCAAAGCAACATCTGGAGCCCTTACA 1119
 DB 1723 GluLeuLeuAspThrTyrValAlaTyrLeuAspGlyGlnAlaPheGlyGlySerSerThr 1742
 QY 1120 GCCGGGACCCCTGGGACCCAGGCGCCAGTCCCTCTCGGGTTCCCAATGCTCGACTTAG 1179
 DB 1743 ProGlyProProProProGlnLeuMetThrProSerLysLeu----- 1757
 QY 1180 CTGAGGCTTGAGAGGTTCACAGCTGGCTGCAGGCGAGAACTTCTCCACTGCCAGGT 1239
 DB 1758 -----ArgGlyArgAlaLeuGlyAlaSerGlyGlyLeuArg-----ProGly 1771
 QY 1240 GAGGACACTTGGTCCCTGACGGGGAGCCAGCTACTTCG-----CATAAGACCTGG 1290

Db 394 ProGlyCysProArgSerAlaArgAsnProGly---CysProArgThrTrpArgArgArg 412
Qy 592 -----CCTAGCAGCCCAAGTAGTCATCACCCTCTCACCCTCA 630
Db 413 SerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThr 432
Qy 631 AGCTATCAACCTGCTACAGCTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGGTT 690
Db 433 GlyGlyArgProAla--AlaPro----- 439
Qy 691 GGGCCAGCTTGGG---CCACTAGCC---GGGCGAGCTGGACAGCGGATGAGCCCTGGCC 744
Db 440 --GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerG 459
Qy 745 CTACCAACCCACATCTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCCCTT 804
Db 459 lYAlaThrProHisProGlu---ArgGlySerGlyProAlaAspPro-ProAlaAlaAla 477
Qy 805 AGGATACAGGACCTGTTCCCTTCGGCGGGGGAAGTCATAGTTAGGGAGCACCTCACA 864
Db 478 ArgLeuPro-----ProGluArgGlnGluProArgLeuProGlnAspLeuAla 493
Qy 865 CTTTCCTCAGAGAGACACACACACCATGACCTCACCTGGGACCCCACTCCAGGCC 924
Db 494 AlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArg 513
Qy 925 CTC-----CAGGCCCAAACTGCTGCCAGCAGCCCTGAAAATGCAAGCTGTGTACGAGTT 978
Db 514 ThrHisArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 533
Qy 979 TGAAGCTAGGAACCCAGGGAAGTACTGTGGT-----CCAGGGAGAGAGCTTGA 1029
Db 534 TrpArg---ArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArg 552
Qy 1030 GGTCTCGACACAGCAGCGGTGGTGGCTGCTGCAAGTAAGCGCGGAGCGGCGCTA 1089
Db 553 ProSerGlyProThr-----GlyGlyAlaAla-----GlyGlyArgProAlaAla 563
Qy 1090 CATTCCAAGCAACATCTCTGGAGCCCTACAGCGGGGACCCCTGGGACCCAGGCGCCAGTC 1149
Db 564 -----ProGlyAlaPro-----GlyThrProAlaAlaProGlyPro--- 575
Qy 1150 ACCCTCTGGGTTCCTCAATGCTTGCAGTTAGCTGAGGCGCTGAAGAGCTCACAGCTGGCT 1209
Db 576 -----GlyGlyGlyAlaAla----- 580
Qy 1210 GCAGGCAGAGAACTTCTCCACTCCCGCTGAGGACACTTGGGTCTCTGAGCGGGAGCCA 1269
Db 581 -----ValProSerGly----- 584
Qy 1270 GCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCAGAGGCGCCACCGAAT 1329
Db 585 AlaThrProHisProGluArgGlySer-----GlyProAlaAsp 597
Qy 1330 CCGTCTCCGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGCACCCAGCTTA 1389
Db 598 Pro-----ProAlaAlaAlaArgLeu 604
Qy 1390 GACACCTCCAGAACACGCCCCCTGATGCAAGATGGCAGATCTGATACCCATTAGAGC 1449
Db 605 ProProGluArgGlnGluProArgLeu----- 613
Qy 1450 CCCGAGAATCTCTCTCTGGATCCAGTTTGCAGCAAAACCCACACCCAGCTCACACAG 1509
Db 614 ProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProProProThrArg----- 631
Qy 1510 CAAACAAATGGACAGCCCGAGGGCTGAA---GCAAAACAGTGTCCCTTGGCTGT 1563
Db 632 -----SerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 646
RESULT 7
SHK1_RAT
ID SHK1_RAT STANDARD; PRT; 2167 AA.

Q9WV48; Q9WU13; Q9WUE8; Q9QZ28;
15-JUN-2002 (Rel. 41, Created)
15-JUN-2002 (Rel. 41, Last sequence update)
15-JUN-2002 (Rel. 41, Last annotation update)
SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
interacting protein) (SSTR interacting protein) (SSTRIP).
SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
DLG4.
RC TISSUE=Brain;
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synamon, a novel neuronal protein interacting with synapse-associated
protein 90/postsynaptic density-95-associated protein.";
RL J. Biol. Chem. 274:27463-27466(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=99360650; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL Neuron 23:569-582(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "The G protein-coupled receptor CL1 interacts directly with proteins
of the shank family.";
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RT "Characterization of the Shank family of synaptic proteins. Multiple
genes, alternative splicing, and differential expression in brain and
development.";
RL J. Biol. Chem. 274:29510-29518(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).
RN [6]
RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
RX PubMed=10433269;
RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
RA Doan A., Akalu V.K., Lanahan A.A., Sheng M., Worley P.F.;
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
postsynaptic density proteins.";
RL Neuron 23:583-592(1999).
RN [7]
RP INTERACTION WITH SPTAN1.
RX PubMed=11509555;
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
multidomain Shank protein family interact with the cytoskeletal
protein alpha-Fodrin.";
RL J. Biol. Chem. 276:40104-40112(2001).


```

Db 331 aAlaAlaAlaProAlaProGlyValLysAlaGluSerProLysArgValValGlnAl 351
Qy 1005 CTGTGGTCCAGGAGAGAGCT---GGAGGTCTTGACACACAGCAAGCGGTGGTGG 1061
Db 331 aAlaProProAlaAlaGlnThrLeuAlaAlaSerGlyProAlaSerThrAlaAlaSer-M 371
Qy 1062 TGAAGATAGCGGAGCGGAGCGGTACATTCACAGC---AACATCTGGAGCCCTTAC 1118
Db 371 etValIleGlyProThrMetGlnGlyAlaLeuProSerProAlaAlaValProProAla 391
Qy 1119 AGCCGGGACCCCT-----GGACCCAGGGCCAGTCACCT 1154
Db 391 laProGlyThrProThrGlyLeuProLysGlyAlaAlaGlyAlaValThrGlnSerLeuS 411
Qy 1155 CTCGGGTTCCAATG-----CTTCGACTTAGCTCGAGCGCTCAAGAGGTCA 1199
Db 411 erArgThrProThrAlaThrThrSerGlyIleArgAlaThrLeuThrProThrValLeuA 431
Qy 1200 CAGACTGGCTCGAGCAGAGAACTTCCTCCACTGCCAGCGTGAGGACACTTGGTCCCTGA 1259
Db 431 laProArgLeuProGlnProProGlnAsnProThrAsnIleGlnAsnPheGlnLeuProp 451
Qy 1260 CGGGAGCCACTACTTCGCATAGACTGGGAGCTACAGATGCTATGTCACAGGAGG 1319
Db 451 roGlyMetValValArgSerGluAsnGlyGlnLeuLeuMetIle---ProGlnGlnA 470
Qy 1320 CCCACGAATCTGTCCCGCTGAGGCTGTCAAGAGTCTGGGATAAGCCCTTAGG 1379
Db 470 la-----LeuAlaGlnMetGlnAlaGlnAlaHisAlaGlnProGlnThrThrMetA 487
Qy 1380 CACAGCTTAGACACCTCCAAAGAACAGGCCCGCTGTGATGATGCGACATCTGATAC 1439
Db 487 laProArgProAlaThrProThrSerAlaPro-----ProValGlnIleSerT 503
Qy 1440 CCATTAGACCCCGAGATTCCTCTCTGGATCCAGTTTCGACCAACCCACACCCCA 1499
Db 503 hrValGlnAlaProGlyThrProIleIle-----AlaArgGlnValThr-----Prot 519
Qy 1500 GCTCACAGCAAAACAATGGAGCGCCAGAGCTGAAGCAACACAGTGTCCCTCTG 1558
Db 519 hrThrIleIleGlnValSerGlnAlaGlnThrThrValGlnProSerAlaThrLeu 538
RESULT 9
YQ36_CAEEL STANDARD; PRT; 963 AA.
AC Q09457;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative cuticle collagen C09G5.6.
GN C09G5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RN Palmer S.;
RA Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
CC LINKS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z46791; CAA86755.1; -.
DR Wormpep; C09G5.6; CE01486.
DR InterPro; IPR002486; Col_cuticle_N.
DR InterPro; IPR000087; Collagen.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF01484; Col_cuticle_N; 1.
KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
KW Multigene family; Collagen.
FT DOMAIN 392..423 TRIPLE-HELICAL REGION.
FT DOMAIN 441..503 TRIPLE-HELICAL REGION.
FT DOMAIN 506..567 TRIPLE-HELICAL REGION.
FT DOMAIN 663..666 POLY-PRO.
FT DOMAIN 685..688 POLY-PRO.
SQ SEQUENCE 963 AA; 107031 MW; AFF895A75909F66E CRC64;

Alignment Scores:
Pred. No.: 0.00117 Length: 963
Score: 185.00 Matches: 141
Percent Similarity: 30.94% Conservative: 44
Best Local Similarity: 23.58% Mismatches: 206
Query Match: 5.92% Indels: 208
DB: 1 Gaps: 31

US-09-762-021A-1 (1-1710) x YQ36_CAEEL (1-963)
Qy 34 TCTGGAGAGAGAGT-----GGAGCAAGACCTCGACTTGGAGGCGCT----- 75
Db 154 SerGlyPheArgAlaProProAlaAlaThrSerThrTyArgProProHisGlySer 173
Qy 76 -----TCAGCCAGG-----CCAGGACAG 93
Db 174 AsnTyrraspAsnTyrrGlyArgGluProAlaSerSerArgArgProTyrrProGlnGln 193
Qy 94 ATGAGGGGGGCTCTATGGAAGCGCGCTCCCTATGGAGCAGGACGCTATCTGGAGCC 153
Db 194 ProProSerThrSerAlaProHisSerSerProAsnAsnArgThrSerLeuTyrrAsnPro 213
Qy 154 GGGATCCCTCCAGAACAGCCCGCCAGGAGGACCCCTAGAGCA-----CAG 198
Db 214 GlnProProLysThrGlyTyrrProThrAsnProArgValProTyrrAsnProGln 233
Qy 199 CCT-----CCACCATCCCT-----AAGGCCCTGCCAGCCACAC 234
Db 234 ProAsnTyrrThrArgGlnProThrTyrrProGluAspAsnArgAlaProTyrrLysProThr 253
Qy 235 CAGTCCCGGAGAACCAAGTGCCTTTACTCTGCCTCCTCCAAAGCGGTCTCTCTCCCGCA 294
Db 254 ArgSerProAsnThrProPro-----ProArg 262
Qy 295 GCACCCAGAGGAGGAGGAGGAGTGTGAACCATGTCTTAAGGACATTGAGCTGTTCAT 354
Db 263 GlnProSerGlyGlyTyrrAspSerAspGlyGlnThrProProSerSerProArgIleTyrr 282
Qy 355 GGGAAAGCTGGAGAGAGCGCCAGGCAAGACAGCAGGAGGAAGAAATTTGGGAAAAAAA 414
Db 283 -----AsnThr 284
Qy 415 CAAGACAGGAGGAGTCTCACCCAGGACAGCAGTACATTGACTGTCTCCAGAGATCAAGTA 474
Db 285 ArgArgProAsnAsnHisGlyProGly-----TyrProGluAspGlnVal 299
Qy 475 CAGCTTCAACCTCTCGGAGAGGCTGGCCACCTGGCTGAAGGAGACAGATGCCCTGAGCT 534
Db 300 ProThrAlaProPro-----ValProGlyGlnGln----- 309
Qy 535 CGTACACATCTCTCAAGTCCCTGAACCTTCATCTCTGCCAGGTGCCCTGAGGCTGGCT 594
Db 310 ArgValProProThrGlnThrArgAsn---ProProAsnProThrAsnThrArgGlnPro 328
Qy 595 AGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTTAAAGCTATCAACCTGTACAGTCTGT 654

```


Db	637	ProArgGlyGlyProGlyGluArgGlyGluInGlyProProGlyGlyProAlaGlyPhePro	655
Qy	640	CCTGCTACAGTCCTGTTCTAAAGCCACCTGAGACTAACCCTTTGGATGGGGTGG	691
Db	657	GlyAlaProGlyGlnAsn-GlyGluProGlyAlaLysGlyGluArgGlyAlaProGlyGly	676
Qy	692	-----GGCCAGCCTGGACCACTAGCCGGGGCGGACTGCAGAGCGATGAGCC	738
Db	676	uLysGlyGluGlyPro-ProGly---AlaAlaGlyProAlaGlyGlySerGlyProA	695
Qy	739	CCTGCCCTACCACCCACATTCTCAGATGACTGCGCACTTCCAGAGCCCT-	788
Db	695	laGlyProProGlyProGlnGly---valLysGlyGluArgGlySerProGlyGlyProG	714
Qy	789	-----CCAGCCACAGCACCTCTAGGATA	810
Db	714	lyAlaAlaGlyPheProGlyGlyArgGlyProProGlyPro-ProGlySerAsnGlyAsn	733
Qy	811	CCAGGACCCGTGTTCCCT---TCGGCGGGGAAGTCATAGTTAGGAGCACCTCACATT	867
Db	734	ProGlyPro---ProGlySerSerGly-	741
Qy	868	TCCTCAGGAGAACACACAAACCATGACCTCAGCCTGCGGGACCCCACTCCAGGCCCT	927
Db	742	AlaProGlyLysAspGlyProProGlyProProGly-	753
Qy	928	CAGCCCCAAACCTGCCACAGCCAGCCCTGAAATGCAAGTCTGTACGAGTTTGAAGCTAG	987
Db	754	--SerAsnGlyAlaProGlySerProGlyLysSerGly-	765
Qy	988	GAACCCACGGGAAGTACTGTGTCTCCAGGAGAGAGCTGGAGTTCTTGGACCCACAGCAA	1047
Db	766	---ProlysGly---AspSerGlyProProGlyGluArgGlyAlaProGlyProGlnGly	783
Qy	1048	CGGTGTGTGCTGGTGAAGAATGAGCGGGACCGGCGCTACATTCCCAAGCAACATCT	1107
Db	784	Pro-----Pro	785
Qy	1108	GGAGCCCT-----ACAGCCGGGAC	1127
Db	786	GlyAlaProGlyProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyPro	805
Qy	1129	CCC-----TGGGACCCAGGACCCAGTCACCCCTCTCGGTTTCCATGCT	1177
Db	806	ProGlyMetProGlyAlaArgGlySerProGlyProGlnGlyIleLysGlyGluAsnGly	825
Qy	1171	TCGACTTAGCTCAGGGCTCAGAGGTCACAGACTGGCTGCAGGCAGACAA-----	1227
Db	826	LysProGlyProSerGlyGlnAsnGlyGluArgGlyProProGlyProGlnGlyLeuPro	845
Qy	1222	--CTTCTCCACTGCCACGGTGAG-----GACACTTGGGTC	1257
Db	846	GlyLeuAlaGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyLeu	865
Qy	1255	CCT-----GACGGGGAGCCAGCTACT-----TCGCATAGACCTGGGA	1297
Db	866	ProGlyArgAspGlyAlaProGlyAlaLysGlyAspArgGlyGluAsnGlySerProGly	885
Qy	1294	GCTACAGATGCTATGCCAGAGGCCCCACGAATCTGTCTCCGCTGGAGGCTGCTAG	1357
Db	886	AlaProGlyAlaProGlyHisProGlyProGlyProValGlyProAlaGlyLysSer	905
Qy	1354	AAGATGCTGGGATAG-----CCCTTAGGCCACCACTTAGCACCTCCCAAG	1407
Db	906	GlyAspArgGlyGluThrGlyProAlaGlyProSerGlyAlaProGlyProAlaGlySer	925
Qy	1402	AACACAGGCCCGCTGATGCAAGATGGCAGATCTGTACCATTTAGAGCCCGCAATTTCC	1467
Db	926	ArgGlyProProGlyProGlnGlyProArgGlyAspLysGlyGluThrGlyGluArgGly	945
Qy	1462	TCATTCTGGATCCCAAG-----TTTGCAGCAAAACCC-----ACACCC	1497
Db	946	AlaMetGlyIleLysGlyHisArgGlyPheProGlyAsnProGlyAlaProGlySerPro	965

```
QY 1498 CAGCTCACACAGCAAAACAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCTC----- 1551
D 966 -----GlyProAlaGlyHisGlnGlyAlaValGlySer 976
QY 1552 -----CCTTCTGGCTGTGTGGAGGCTCC 1575
D 977 ProGlyProAlaGlyProArgGlyProValGlyProSer 989

RESULT 11
DRPL_RAT
ID DRPL_RAT STANDARD: PRT: 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.F., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidoluysian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -!- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -!- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31777; AAA80337.1; -
DR EMBL: X89453; CAA61623.1; -
DR InterPro: IPR002951; Atrophin.
DR Pfam: PF03154; Atrophin-1; 2.
DR PRINTS: PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT DOMAIN 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
```

```
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCA9F9B1F CRC64;
Alignment Scores:
Pred. No.: 0.00125 Length: 1183
Score: 184.50 Matches: 154
Percent Similarity: 32.32% Conservative: 49
Best Local Similarity: 24.52% Mismatches: 196
Query Match: 5.90% Indels: 229
DB: 1 Gaps: 34

US-09-762-021a-1 (1-1710) x DRPL_RAT (1-1183)
QY 34 TCTGGAGGAAGAGCTGGAGCAAAACACCTCGACTTGGAGGCCCTTCAGCAGGCCAGCCAGGACAG 93
D 265 SerSerGlyAlaSerGlyAla-----ProProAlaLysProProAsn 278
QY 94 ATGGAGGGGGCTGCTATGTGGAAGCCCGCTCCCTATGAGCAGGCGACCTATCTCGAGCC 153
D 279 -----ThrProValGlyAlaGlyAsnLeuProSerAla 289
QY 154 GGGATCCCTCCAGAAACAGCCACAGGACCCCTAGAGCAGACAGCTCCACCATCCCC 213
D 290 ProProProAla-----ThrPheProHisValThrProAsnLeuProPro-ProProAlaLe 308
QY 214 AAGGCCCTGCCACGCCACACAGTGGCCCGAGAACCAAGTGCCTTTACTTCTGCTCTCC 273
D 308 uArgProLeu---AsnAsnAlaSerAlaSerProProGlyMetGlyAlaGlnProIlePr 327
QY 274 AAGGC-----GGTCTCTTCCCGGAGGA 297
D 327 o-GlyHisLeuProSerProHisAlaMetGlyGlnGlyMetSerGlyLeuProProGlyP 347
QY 298 CCCAGAGGAGGAGGAGGAAGTCTGAACCATGCTCTAAAGGGACATTTAGCTGTTCATGGG 357
D 347 roGluLysGlyPro-----ThrLeuAlaProSer---- 356
QY 358 AAGCTGGAGGAAGGCCCGCAGGCAAGACAGCAGGAAGAAATTTGGGAAAAAACA 417
D 357 -----ProHisProLeuProProAlaSerSerSerAlaProGlyProProM 372
QY 418 GCACCGAGGAGGTCTCACCCAGGCACATGACATTCACCTTCCAGAGATCAAGTACAG 477
D 372 eArgTyProTySerSerSerCysSerSerSerValAlaAlaSerSerSerSerA 392
QY 478 CTTCACCTCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCC----- 527
D 392 laAlaThrSer-----GlnTyProAlaSerGlnT 402
QY 528 --CTGAGCTGTACACATCTCTTCAAGTCCCTGAACTTCATCTCGGCCAGGTGCCCTGA 585
D 402 hrLeuProSerTyProHisSerPheProPro-----ProThrSerMets 417
QY 586 GGCTGGCCTAGCAG---CCCAGTGTATCTACCCCTCTCACCCCTCACCCCTATCAACCT 642
D 417 erValSerAsnGlnProProLysTyThrGlnProSerLeuProSerGlnAlaVal---- 435
QY 643 GCTACAGCTCTCTTAAGCCACCTGAGAGTAACCTTTGGATGGGTGGGCCCCAG---- 698
D 436 -----TrpSerGlnGlyP 440
QY 699 -----CCTGGAC 705
D 440 roProProProProProTyGlyArgLeuLeuProAsnAsnAsnThrHisProGlyP 460
QY 706 CACTAGCCGGCCGACTGA-----CAGGCCATGAGCCCTGCTTACCAACCCAC 756
D 460 roPhePro---ProThrGlyGlyGlnSerThrAlaHisProProAlaHisHisH 479
QY 757 ATTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCAA----- 796
D 479 isHisGlnGlnGln--GlnGlnProGlnProGlnProGlnHisHisHisG 498
QY 797 -----GCACCTTAGGATACAGGACCTGTTCCTTCGGCGGGGAAG 840
```

```
Db 498 yAsnSerGlyProProGlyAlaTyrProHisProLeuGlu----- 513
QY 841 TCATAGTTAGGAGCACCCTCAGCTCTTCTCAGGAGAAGACACAAACCATGACCTCA 900
Db 514 -----SerSerAsnSerHis-----HisAlaHisProTyrAsnMetSerPro-S 528
QY 901 GCTCGGGGACCAACTCCAGGCCCTCCAGCCCCA---AACCTGCCAGCCAGCCCTGAA 957
Db 528 erLeuGlySerLeuArgProTyrProGlyProAlaHisLeuPro----- 544
QY 958 ATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCAGGGGAGCTGCTGCTCCAGG 1017
Db 545 -----SerHisGlyGln-----ValSerTyrS 552
QY 1018 AGAAGAGCTGGAGTCTTGAGACACAGCAAGCGGTGGTGGTGAAGATGAGCGGG 1077
Db 552 er-GlnAlaGlyProAsnGlyProProValSerSerSerAsnSerSerGlySerSer 571
QY 1078 ACGGAGCGGCTACATTCACAAGCAACATCCT-----GGAGCCCTACAGCCGGG 1125
Db 572 SerGlnAlaAlaTyrSerCysSerHisProSerSerSerGlnGlyProGlnGlyAlaSer 591
QY 1126 GACCCCTGGGACCCA---GGGCCATCACCCTCTCGGGTTCCAATGCTTCGACTAGCTC 1182
Db 592 TyrProPheProProValProProIleThrThrSerSerAlaThrLeuSerThrValIle 611
QY 1183 -----GAGCCTGAAGAGTGCACAGACTGGCTGCAGGCAGAGAACTTCTCCAC 1230
Db 612 AlaThrValAlaSerSerProAlaGlyTyrIleThrAlaSer----- 625
QY 1231 TCCACGCTGAGGACACTTGGTCCCTGCAGGGGAGCCA----- 1269
Db 626 -----ProGlyProProGlnTyrSerLysArgAlaPro 637
QY 1270 ---GCTACTTCCATAAGACCTGG-----GGAGCTACAGATGCT 1305
Db 638 SerProGlySerTyrIleThrAlaThrProGlyTyrLysProGlySerProSer 657
QY 1306 ATGTCACAGGAGGCC-----ACGAATCCTGTCCCGCTGGAGGCTGT 1350
Db 658 PheArgThrGlyThrProProGlyTyrArgGlyThrSerPro---ProAlaGlyProGly 676
QY 1351 CAGAAAGTGTCTGGGATAAGCCCTTAGCCACCACTTAGACACCTCCAAAGAACAGGCC 1410
Db 677 ThrPheLysProGlySerProThrValGly-----ProGly 688
QY 1411 CCGCTGATGCAAGATGGCAGATCTGATACCCATTAGACCCCGGAGAAATCTCTTCTGGA 1470
Db 689 ProLeuProProAlaGly-----ProSerSerLeuSerSerLeu 701
QY 1471 TCCAGTTTGCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGGACAGGCCA 1530
Db 702 ProProProProAlaAlaProThr-----ThrGlyPro 712
QY 1531 GAGGCTGAAGCAACAGTGTCCCTCTGCTGTGTGGAGCCTCCCGACCAACCACTAT 1590
Db 713 ProLeuThrAlaThrGlnIleLysGln-----GluProAlaGluGluTyr 727
QY 1591 TTATTTTACCTTTTCCCAACCTGGAGCATTTATGCTTAGCTGCTCAAGAATCTGTC 1650
Db 728 GluThrProGluSerProValProProAlaArgSerPro----- 740
QY 1651 AGTCCTCTCTCT 1662
Db 741 SerProProPro 744
RESULT 12
CA17_HUMAN
ID CA17 HUMAN STANDARD; PRT; 2944 AA.
AC Q02388; Q14054; Q16507;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
```

```
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC
GN COL7A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eumalia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94327588; PubMed=8051117;
RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen. Complete primary sequence of the
RT alpha 1(VII) chain and identification of intragenic polymorphisms.";
RL J. Biol. Chem. 269:20256-20262(1994).
RN [2]
RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93338437; PubMed=1307247;
RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,
RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E.,
RA Uitto J.;
RT "The large non-collagenous domain (NC-1) of type VII collagen is
RT amino-terminal and chimeric. Homology to cartilage matrix protein,
RT the type III domains of fibronectin and the A domains of von
RT Willebrand factor.";
RL Hum. Mol. Genet. 1:475-481(1992).
RN [3]
RP SEQUENCE OF 815-1439 FROM N.A.
RX MEDLINE=91334380; PubMed=1871109;
RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W.,
RA Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;
RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
RT gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
RN [4]
RP SEQUENCE OF 369-1255 FROM N.A.
RX MEDLINE=93107742; PubMed=1469284;
RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisayan P.S.,
RA Cook M.E., Wright J., Brigaman R.A., Hunt S.W. III;
RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain
RT adhesion proteins involved in tissue-specific organization of
RT extracellular matrix.";
RL J. Invest. Dermatol. 99:691-696(1992).
RN [5]
RP SEQUENCE OF 340-675 FROM N.A.
PC TISSUE-Keratinocytes;
RX MEDLINE=92231902; PubMed=1567409;
RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
RT "Molecular cloning and characterization of type VII collagen cDNA.";
RL Biochem. Biophys. Res. Commun. 183:958-963(1992).
RN [6]
RP SEQUENCE OF 2395-2944 FROM N.A.
RX MEDLINE=93271985; PubMed=8499916;
RA Greenspan D.S.;
RT "The carboxyl-terminal half of type VII collagen, including the non-
RT collagenous NC-2 domain and intron/exon organization of the
RT corresponding region of the COL7A1 gene.";
RL Hum. Mol. Genet. 2:273-278(1993).
RN [7]
RP SEQUENCE OF 1-87 FROM N.A.
PC TISSUE-Placenta;
RX MEDLINE=94375010; PubMed=8088784;
RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,
RA Uitto J., Greenspan D.S.;
RT "Structural organization of the human type VII collagen gene (COL7A1),
RT composed of more exons than any previously characterized gene.";
RL Genomics 21:169-179(1994).
RN [8]
RP REVIEW ON DEB VARIANTS.
RX MEDLINE=98041696; PubMed=9375848;
RA Jaervikallio A., Pulkkinen L., Uitto J.;
RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in
RT the type VII collagen gene (COL7A1).";
```



```
QY 175 CCACGAGGACCCCTAGACACAGCCCTCCACCACCTCCCAAGGCC---CCTGCCACGCCA 231
DB 1333 ProArgGlyAspProGlyGluArgGlyProArgGlyProGlyGluProGlyAlaPro 1352
QY 232 CACAGTCCCGAGAACCAAGTGCCTTACTCTGCTCCTCCAGGCGGTCTCTCCCC 291
DB 1353 GlyGlnValIleGlyGlyGluGlyProGlyLeuProGlyArgLysGlyAspProGlyPro 1372
QY 292 CGAGGACCCAGAGAGGAGGAGGAGT---CCTGAACCATCTCTCAAGGGACATTGAGCT 348
DB 1373 SerGlyProProGlyProArgGlyProLeuGlyAspProGlyProGlyProGlyPro 1392
QY 349 GTTCATGGAAAGCTGGAGAGGCCAGGCAAGACACAGCAGGAGAGAAATTGGGAA 408
DB 1393 LeuProGlyThrAlaMetLysGlyAspLysGlyAspArgGlyGlu-----Arg 1408
QY 409 AAAAAACAAGACACGAGGAGTCTCAC----- 435
DB 1409 GlyProProGlyProGlyGlyGlyGlyLeuAlaProGlyGluProGlyLeuProGlyLeu 1428
QY 436 CAGGCACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCTCTGGGAAG 495
DB 1429 ProGlySer-----ProGlyProGlyProGlyProGlyProGlyProGlyLys 1443
QY 496 GCTGCCACCTGGCTGAAGAGACAAAGTGCCTGAGCTCTACACATCTCTTCAAGTC 555
DB 1444 LysGlyGluLysGlyAspSerGluAsp-----GlyAla 1454
QY 556 CTGNACTTCATCTGCGCCAGCTGCTGAGGCTGCTGAGGCTGAGCCGACCCAGTGTCTCACC 615
DB 1455 ProGlyLeu---ProGlyGln-----ProGlySerProGlyGlyGlnGly 1468
QY 616 CTCTCTCACCCCTAAAGCTATCAACCTGCTACAGCTCTGCTAAGCCACCTGAGAGTAA 675
DB 1469 ProArgGlyPro-----ProGly----- 1474
QY 676 CCTTTGGATGGGTTGGCCCGAG-----CCTGGACCACTA-----GCC 713
DB 1475 -----AlaIleGlyProLysGlyAspArgGlyPheProGlyProLeuGlyGluAla 1491
QY 714 GGGCGGACTGCACAGGCGATGAGCCCTGCTACCAACCCACATCTCAGATGACTGGC 773
DB 1492 GlyGluLysGlyGluArgGlyProProGlyProAlaGlySerArgGly---LeuProGly 1510
QY 774 AACTTCCAGACCCCTCCAGCAACACCCCTTAGGATACAGGACCTGTTTCCCTCGCTGGC 833
DB 1511 ValAlaGlyArgProGlyAlaLysGlyPro-Glu-----GlyProProGlyProThrGly 1528
QY 834 GGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAAGACACACAACCA-- 891
DB 1528 yArgGlnGly-----GluLysGlyGluProGlyArgProGly 1540
QY 892 -TGACCTCAGCTGGGACCCCAACTC-----CAGGCCCTCCAGCCCAAACTG 941
DB 1540 yAspProAlaValValGlyProAlaValAlaGlyProLysGlyGlyAspValGly 1560
QY 942 CCCAGCCAGCCCTGAATGCAAGTCTGTACAGAGTTTGAAGCTAGGAACCCAGGGAAC 1001
DB 1560 yProAlaGlyProArgGlyAlaThrGlyValGln----- 1571
QY 1002 TGACTGTGTCAGGAGAGAAGCTGGAGGTTCTGGACCAAGCAGCGGTGGTGGCTGG 1061
DB 1572 -----GlyGluArgGlyProProGlyLeuValLeuProGly 1583
QY 1062 TGAAGATGAGGGGCGGAGCGGCTACATTCCAGCAACATCTCGGAGCCCTACAGC 1121
DB 1583 yAsp-----ProGly----- 1588
QY 1122 CGGGACCCCTGGACCCAGGCGGAGTCAACCTCTCGGGTTTCCAATGCTTGGACTAGCT 1181
DB 1588 sGlyAspProGlyAspArgGlyProIleGlyLeuThrGlyArgAlaGlyProProGlyAs 1608
QY 1182 CGAGGCCTGAAGAGGTCACAGACTGGCTGCAGGCAGAGAA-----CTTCTCCA 1229
```

```
DB 1608 pSerGlyProProGlyGluLysGlyAspProGlyArgProGlyProGlyProValGly 1628
QY 1230 CTGCCAGCGGTGAGCACACTGGGTC-----CCTGACGGGGAGCC 1268
DB 1628 yProArgGlyArgAspGlyGluValGlyGluLysGlyAspGlyProProGlyAspPr 1648
QY 1269 AGCTACTTCGATTAAGACCTGGGAGCTACAGATGCTATGCTCCACAGGAGGCCACGAA 1328
DB 1648 oGlyLeuProGlyLysAla---GlyGluArgGlyLeuArgGlyAlaProGlyValArgGly 1667
QY 1329 TCTGTCTCCCGCTGAGGCTGTCAAGAGGATGCTGGGATAAGCCCTTAGGCACACAGCTT 1388
DB 1667 yProValGlyGluLysGlyAspGlnGlyAspProGly----- 1679
QY 1389 AGACACCTCCAGAACCCAGGCCCGCTGATGTCAGATGCGCAGATCTGTATCCCATTTAGAG 1448
DB 1680 -----GluAspGlyArgAsnGly-----Se 1686
QY 1449 CCCCAGAAATTC-----TCTTCTGGATCCAGTTTCAGCAAAACCCACACC 1496
DB 1686 rProGlySerSerGlyProLysGlyAspArgGlyGluProGlyProGlyProProGly 1706
QY 1497 CCAGCTCACACAGCAAAACAATGGACAGCCCGGAGCTGAAGCAACAGCTGTCCCT 1554
DB 1706 yArgLeuValAsp-----ThrGlyProGlyAlaArgGluLysGlyGluPro 1721

RESULT 13
MAPA_RAT
ID MAPA_RAT STANDARD; PRT; 2774 AA.
AC P34926;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain
DE LC2].
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92355629; PubMed=1379599;
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT one messenger RNA.";
RL J. Biol. Chem. 267:16561-16566(1992).
CC -|- FUNCTION: Structural protein involved in the filamentous cross-
CC bridging between microtubules and other skeletal elements.
CC -|- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -|- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -|- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC THEIR MORPHOLOGY.
CC -|- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -|- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -|- PFM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -|- SIMILARITY: TO MAP1B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```



```
Db 2277 ProAlaProSerAlaSerLeuAspLeuAlaProAlaProAlaProAlaProAlaPro 2295
      |||
Qy 1339 GCTGGAGGCTCAGAGGATGCTGGGAT-----AAGCCCTTAGGCACC 1383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2296 AlaProGlyLeuProGlyAspLeuGlyAspGlyThrLeuProCysArgProGluCysThr 2315
Qy 1384 AGCTTAGACACTCCAGAACAGCCGCCGCTGATGCAGATGCGAGATCTGATACCCAT 1443
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2316 Gly---GluLeuThrLysLysProSerProPheLeuSerPro----- 2328
Qy 1444 TAGAGCCCGGAGATTCCTCTCTGGA---TCCAGATTGAGCAAAACCCACACCCAG 1500
      ||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 2329 ---SerGlyAspHisGluAlaAsnGlyProGlyGluThrSerLeuAsnProGlyPhe 2347
Qy 1501 CTCACA-----CAGCAAAACATGGACGCCAGAGGCT---GAAGCAAAACAGT 1548
      : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2348 ValThrAlaThrAlaGluLysGluGluAlaGluAlaProHisAlaTrpGluArgGlySer 2367
Qy 1549 GTCCTCTGCTGTGTGGAGCCTCCACGTAACCACTATTATTATTACCTCTTTCCC 1608
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2368 TrpProGluGlyAlaGluArgSerSerArgProAspThrLeuLeuSerSerGluGlnPro 2387
Qy 1609 ---AAACTGCA 1617
      : : : : : |||
Db 2388 LeuArgProGly 2391
      : : : : : |||
RESULT 14
CALL_CHICK
ID CALL_CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-153 FROM N.A.
RX MEDLINE=88056316; PubMed=3678834;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RT of the chicken pro alpha 1(I) collagen mRNA.";
RL Gene 56:71-78(1987).
RN [2]
RP SEQUENCE OF 1-144 FROM N.A.
RX MEDLINE=88007542; PubMed=2820966;
RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RT "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP SEQUENCE OF 152-1187.
RX MEDLINE=82231995; PubMed=7093229;
RA Highbarger J.H., Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Kang A.H., Gross J.;
RT "Amino acid sequence of chick skin collagen alpha 1(I)-C8 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP SEQUENCE OF 1200-1205.
RX MEDLINE=72243016; PubMed=5047697;
RA Eyre D.R., Glimcher M.J.;
RT "Evidence for a previously undetected sequence at the carboxyterminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP SEQUENCE OF 981-1453 FROM N.A.
```

```

RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RT including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP SEQUENCE OF 1311-1453 FROM N.A.
RX MEDLINE=80134546; PubMed=6987088;
RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,
RA Pastan I., Decrombrughe B., Fietzek P.P., Olsen B.R.;
RT "Nucleotide sequence of a collagen cDNA-fragment coding for the
RT carboxyl end of pro alpha 1(I)-chains.";
RL FEBS Lett. 111:61-65(1980).
CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
      (FIBRILLAR FORMING COLLAGEN).
CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
      BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
      HYDROXYAPATITE.
CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
      UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -!- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M17839; AAA48704.1; -
DR EMBL; M17838; AAA48704.1; JOINED.
DR EMBL; V00401; CNA23695.1; -
DR EMBL; M10571; AAA48671.1; ALT_SEQ.
DR EMBL; M17607; AAA48672.1; -
DR PIR; A02857; CGCH15.
DR PIR; A27179; A27179.
DR PIR; A29367; A29367.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib-collagen_C.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF00093; vwc; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Collagen; 2.
DR ProDom; PD002078; Fib-collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1205 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1206 1453 C-TERMINAL PROPEPTIDE.
FT DOMAIN 31 89 VWFC.
FT MOD_RES 152 152 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 254 254 HYDROXYLATION (POTENTIAL).
FT MOD_RES 851 851 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1081 1081 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1097 1097 HYDROXYLATION (POTENTIAL).
FT MOD_RES 1153 1153 HYDROXYLATION (ONLY 3-HYDROXYPRO AND THE
      ONLY HYDROXYLATED PRO IN POSITION X (IN
      THE G-X-Y UNIT IN THE ALPHA 1(I) CHAIN)).
      F -> L (IN REF. 5).
      O -> H (IN REF. 6).
FT CONFLICT 1187 1187
FT CONFLICT 1441 1441
SQ SEQUENCE 1453 AA; 137789 MW; 3BC6152134271F4D CRC64;
-----
Alignment Scores: 0.00199 Length: 1453
Pred. No.: 181.00 Matches: 140
Score:
```

Percent Similarity: 30.36% Conservative: 30
 Best Local Similarity: 25.00% Mismatches: 201
 Query Match: 5.79% Indels: 189
 DB: 1 Gaps: 31

US-09-762-021A-1 (1-1710) x CALL_CHICK (1-1453)

QY 25 GCAGAGGCTCTGGAGAGAGCTGGAGCAAGAGCTCGACTTGG----- 69
 DB 742 AlaAspGlyAlaProGlyLysAspGlyLeuArgGlyLeuThrGlyProIleGlyProPro 761
 QY 70 AGCCCTTCACGACGAGCCAGACATG-----GAGGGGCTCTGTATGGAAGCCGCTCCC 126
 DB 762 GlyProAlaGlyAlaProGlyLysGlyGlyAlaGlyProGlyProAlaGlyPro 781
 QY 127 TATGGAGCA-----GGCAGCTATCT 147
 DB 782 ThrGlyAlaArgGlyAlaProGlyAspArgGlyGluProGlyProGlyProAlaGly 801
 QY 148 GGAGCGGGGATCCCTCCAGAACAGCCCA-----GAGGACCT 189
 DB 802 PheAlaGlyProGlyAlaAspGlyGlnProGlyAlaLysGlyLeuThrGlyAspAla 821
 QY 190 AGACACAGCCTCCACCATCCCAAGGCCCTGCGCCACGCCACACAGTGGCCGAGAAC 249
 DB 822 GlyAlaLysGlyAspAlaGlyProGlyProAlaGlyProThrGlyAlaPro----- 839
 QY 250 AAGTGCCTTACTCTCCCTCCCAAGGGCGTCTTCCCGAGGACCCAGAGGGA 309
 DB 840 -----GlyProAlaGly***ValGlyAlaPro---GlyProLysGlyAla 853
 QY 310 CGAGGAAGTCTGAACCATCTCTTAAGGACATTTGAGCTGTTTCATGGGAAAGCTGGAGAA 369
 DB 854 ArgGlySerAlaGly-----ProProGlyAlaThrGlyPheProGlyAlaAlaGly--- 870
 QY 370 GGCCAGGCAAGACAGCAGGAAGAANAATTGGGAANAACACAGGACCCAGGAGG 429
 DB 871 -----ArgValGlyPro----- 874
 QY 430 TCTCACCCAGGACAGTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAA---CCT 486
 DB 875 -----ProGly-----ProSerGlyAsnIleGlyLeuProGlyPro 886
 QY 487 CTGTGGAGGCTGGCCACCTGCTGAAGGAGACAAGTGCCTGAGCTGCTACATCCT 546
 DB 887 ProAlaGly-----Lys***GlySerLysGlyProArgGlyGluThrGlyPro 904
 QY 547 CTTCAAGTCCCTGAACCTTCCTGCGCCAGTGCCTGAGGCTGGCTAGACGCCAAGT 606
 DB 905 Ala-----GlyArgProGlyGluProGly 912
 QY 607 GATCTCACCCCTCCACCCCTAAAGCTATCAACCTGCTACAGTCTGCTTAAGCCACC 666
 DB 913 ProAlaGlyProGlyPro----- 919
 QY 667 TGAGAGTAACCTTTGATGGGTGGGCCCCAGCCCTGGACCACTAGCGGGCGACTGGAC 726
 DB 920 -----ProGlyGluLysGlySerProGlyAlaAspGlyProIleGly----- 933
 QY 727 AGGCGATGAGCCCTCCCTCCACACCCACATTCAGATGACTGGCAACTTCCAGAGCC 786
 DB 934 -----AlaProGlyThrProGly-----ProGlnGlyIleAla 944
 QY 787 CTCACGCAAGCACCTCTAGGATACACCA-----CCCTGTTTC 825
 DB 945 GlyGlnArgGlyValValGlyLeuProGlyGlnArgGlyGluArgGlyPheProGlyLeu 964
 QY 826 CCTTCGGCGGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAACACACA 885
 DB 965 ProGlyProSerGlyGluProGlyLysGlnGlyProSerGlyAlaSerGlyGluArgGly 984
 QY 886 CAACCATGACCTCAGCCTGGGAGCCCACTCCAGCCCTCCAGCCCAACCTGCCCCA 945

DB 985 ProProGlyPro-----MetGlyProProGlyLeuAlaGlyProProGly-----Glu 1000
 QY 946 GCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGAAGTAC 1005
 DB 1001 AlaGlyArgGlyAla-----ProGlyAlaGlu 1010
 QY 1006 TGTGTCTCCAGGAGAGCTGGAGTCTTGACACACAGCAAGCGTGGTGGTGGTAA 1065
 DB 1011 GlyAlaProGlyArgAsp---GlyAlaAlaGlyProLys----- 1022
 QY 1066 GAATGAGGGGACGAGCGGCTACATTCCCAAGCAACATCTCTGGAGCCCTACACCGGG 1125
 DB 1023 -----GlyAspArgGlyGluThrGlyProAlaGlyProProGlyAlaPro-----Gly 1038
 QY 1126 GACCCTGGGACCCAGGCGCTACCTCTCGGGTTCATGCTTCGACTTAGCTCGAG 1185
 DB 1039 AlaProGlyAlaProGlyProValGlyProAlaGlyLysAsnGlyAsp----- 1054
 QY 1186 GCCTGAAGAGGTACAGACTGGCTGCAGGCGAGAGAACTTCTCCACTGCCACGGTGAGGAC 1245
 DB 1055 -----ArgGlyGluThrGlyProAlaGly----- 1062
 QY 1246 ACTTGGTCCCTGAGGGGAGCCAGCTACTTCGCATAGAGCTGGGAGCTACAGTCT 1305
 DB 1063 -----ProAlaGlyProPro-----GlyProAlaGlyAla 1072
 QY 1306 ATGTCCACAGGAGGACCCACGAACTCTGTCCCGGCTGGAGGCTGCAGAAAGTGTGG 1365
 DB 1073 ArgGlyProAlaGlyProGlnGlyProArgGlyAspLysGlyLeuThrGlyGlnGly 1092
 QY 1366 GATAAGCCTTAGGACCCAGCTTACAGCTCCAGAACAGCCCGCTGTATGCAAGAT 1425
 DB 1093 AspArgGlyMetLysGlyHisArgGlyPheSerGlyLeuGlnGlyProProGlyPro 1112
 QY 1426 GCCACATCTGATACCCATAGAGCCCC---GAGAACTCTCTTCTGGATCCAGTTGCA 1482
 DB 1113 Gly-----AlaProGlyGluGlnGlyProSerGlyAlaSerGlyPro 1126
 QY 1483 GCAAAACCCACACCCAGCTCACAGCAAAACAAATGCAGAGCCAGAGGCTGAAGCA 1542
 DB 1127 AlaGlyPro-----ArgGlyProProGlySerAla 1136
 QY 1543 -----AACAGTGTCCCTTGTGGTGTGGAGCCCTCCC 1576
 DB 1137 GlyAlaAlaGlyLysAspGlyLeuAsnGlyLeuPro---GlyProIleGly-ProPro 1154

RESULT 15

CALL_MOUSE STANDARD: PRT: 1453 AA.

AC P11087; Q50635;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1 OR COL1A1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N;
 RA Li S.W., Khillan J., Prockop D.J.;
 RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
 of type I procollagen.";
 RL Matrix Biol. 14:593-595(1995).
 RN [2]
 RP SEQUENCE OF 518-1128 FROM N.A.
 RC MEDLINE=86137403; PubMed=3841523;
 RA French B.T., Lee W.-H., Maul G.G.;
 RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
 collagen protein.";

Gene 39:311-312(1985).
[3]
SEQUENCE OF 735-1130 FROM N.A.
MEDLINE=83141374; PubMed=6298597;
Monson J.M., Friedman J., McCarthy B.J.;
"DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene.";
Mol. Cell. Biol. 2:1362-1371(1982).
[4]
SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
MEDLINE=831517109; PubMed=6219867;
Monson J.M., McCarthy B.J.;
"Identification of a half/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
DNA 1:59-69(1981).
[5]
SEQUENCE OF 1442-1453 FROM N.A.
MEDLINE=88124276; PubMed=3340560;
Moolenaar K., Harbers K.;
"Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
Nucleic Acids Res. 16:773-773(1988).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
(FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; U08020; AAA88912.1; -
DR EMBL; X15896; CAA33904.1; -
DR EMBL; M14423; AAA37333.1; -
DR EMBL; M17491; AAA37334.1; -
DR EMBL; X06753; CAA29927.1; -
DR EMBL; K03036; AAA37332.1; -
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; A23982; A23982.
DR MGI; MGI:88467; Colla1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib.collagen_C.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWFC.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT

FT	DOMAIN	168	1181	TRIPLE-HELICAL REGION.
FT	DOMAIN	1182	1207	NONHELICAL REGION (C-TERMINAL).
FT	CARBOHYD	56	56	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1354	1354	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	734	736	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1082	1084	CELL ATTACHMENT SITE (POTENTIAL).
FT	CONFLICT	1450	1450	A -> V (IN REF. 5).
SQ	SEQUENCE	1453 AA;	137944 MW;	3B802E535DF01808 CRC64;

Alignment Scores:
Pred. No.: 0.00227 Length: 1453
Score: 180.00 Matches: 147
Percent Similarity: 32.93% Conservative: 42
Best Local Similarity: 25.61% Mismatches: 226
Query Match: 5.76% Indels: 159
DB: 1 Gaps: 35

US-09-762-021A-1 (1-1710) x CALL_MOUSE (1-1453)

Qy	1	GGCAGACGACTGAAGACACCGCTGCA	-----GAAGGCTCTGGAGGAAGA	45
Db <td>618<td>GlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyPro</td><td> </td><td>637</td></td>	618 <td>GlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyPro</td> <td> </td> <td>637</td>	GlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyPro		637
Qy <td>46<td>GCTGGAGCAAGACCTCGACTTGGAGCGCTTACGACGAG</td><td>-----CCAGGACAGATGGAG</td><td>99</td></td>	46 <td>GCTGGAGCAAGACCTCGACTTGGAGCGCTTACGACGAG</td> <td>-----CCAGGACAGATGGAG</td> <td>99</td>	GCTGGAGCAAGACCTCGACTTGGAGCGCTTACGACGAG	-----CCAGGACAGATGGAG	99
Db <td>638<td>AlaGlyProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLeu</td><td> </td><td>656</td></td>	638 <td>AlaGlyProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLeu</td> <td> </td> <td>656</td>	AlaGlyProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLeu		656
Qy <td>100<td>GGGCGCTGCTATGAAAGCGCGCTCCCTATGAGCA</td><td>-----</td><td>135</td></td>	100 <td>GGGCGCTGCTATGAAAGCGCGCTCCCTATGAGCA</td> <td>-----</td> <td>135</td>	GGGCGCTGCTATGAAAGCGCGCTCCCTATGAGCA	-----	135
Db <td>657<td>-----GlyAlaProGlyProSerGlyAlaArgGlyGluArgGlyPheProGly</td><td> </td><td>672</td></td>	657 <td>-----GlyAlaProGlyProSerGlyAlaArgGlyGluArgGlyPheProGly</td> <td> </td> <td>672</td>	-----GlyAlaProGlyProSerGlyAlaArgGlyGluArgGlyPheProGly		672
Qy <td>136<td>-----GGCAGCGCTATCTGGAGCGCGGATCCCTCCAGAACGCCACCA</td><td> </td><td>180</td></td>	136 <td>-----GGCAGCGCTATCTGGAGCGCGGATCCCTCCAGAACGCCACCA</td> <td> </td> <td>180</td>	-----GGCAGCGCTATCTGGAGCGCGGATCCCTCCAGAACGCCACCA		180
Db <td>673<td>GluArgGlyValGlnGlyProGlyProAlaGlyProArgGlyAsnGlyAlaPro</td><td> </td><td>692</td></td>	673 <td>GluArgGlyValGlnGlyProGlyProAlaGlyProArgGlyAsnGlyAlaPro</td> <td> </td> <td>692</td>	GluArgGlyValGlnGlyProGlyProAlaGlyProArgGlyAsnGlyAlaPro		692
Qy <td>181<td>GAGGACCTAGAGCAGCAGCTCCACC</td><td>-----ATCCCCAAGGCCCTGCCAGCCACACAG</td><td>237</td></td>	181 <td>GAGGACCTAGAGCAGCAGCTCCACC</td> <td>-----ATCCCCAAGGCCCTGCCAGCCACACAG</td> <td>237</td>	GAGGACCTAGAGCAGCAGCTCCACC	-----ATCCCCAAGGCCCTGCCAGCCACACAG	237
Db <td>693<td>GlyAsnAspGlyAlaLysGlyAspThrGlyAlaProGlyAlaProGlySer</td><td> </td><td>711</td></td>	693 <td>GlyAsnAspGlyAlaLysGlyAspThrGlyAlaProGlyAlaProGlySer</td> <td> </td> <td>711</td>	GlyAsnAspGlyAlaLysGlyAspThrGlyAlaProGlyAlaProGlySer		711
Qy <td>238<td>TGCCCGAGAACCAAGTGCCTTACTCTGCCTCCCAAGGGGCTCTTCCCCCGAGGA</td><td>-----</td><td>297</td></td>	238 <td>TGCCCGAGAACCAAGTGCCTTACTCTGCCTCCCAAGGGGCTCTTCCCCCGAGGA</td> <td>-----</td> <td>297</td>	TGCCCGAGAACCAAGTGCCTTACTCTGCCTCCCAAGGGGCTCTTCCCCCGAGGA	-----	297
Db <td>712<td>AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeu</td><td> </td><td>729</td></td>	712 <td>AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeu</td> <td> </td> <td>729</td>	AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeu		729
Qy <td>298<td>CCCAGAGGAGGACGAGGAAGTGTGAACCATGCTCTAAGGACACATTGAGCTGTTTCATGGG</td><td> </td><td>357</td></td>	298 <td>CCCAGAGGAGGACGAGGAAGTGTGAACCATGCTCTAAGGACACATTGAGCTGTTTCATGGG</td> <td> </td> <td>357</td>	CCCAGAGGAGGACGAGGAAGTGTGAACCATGCTCTAAGGACACATTGAGCTGTTTCATGGG		357
Db <td>730<td>ProLysGlyAspArgGlyAspAlaGly</td><td>-----ProLysGly</td><td>741</td></td>	730 <td>ProLysGlyAspArgGlyAspAlaGly</td> <td>-----ProLysGly</td> <td>741</td>	ProLysGlyAspArgGlyAspAlaGly	-----ProLysGly	741
Qy <td>358<td>AAAGCTGGAGAGGCCGCCAGGCAAGACAGCAGGAGGAGAAATTTGGGAAAAAAACAA</td><td> </td><td>417</td></td>	358 <td>AAAGCTGGAGAGGCCGCCAGGCAAGACAGCAGGAGGAGAAATTTGGGAAAAAAACAA</td> <td> </td> <td>417</td>	AAAGCTGGAGAGGCCGCCAGGCAAGACAGCAGGAGGAGAAATTTGGGAAAAAAACAA		417
Db <td>742<td>---AlaAspGlySerProGlyLysAspGlyAlaArg</td><td>-----</td><td>752</td></td>	742 <td>---AlaAspGlySerProGlyLysAspGlyAlaArg</td> <td>-----</td> <td>752</td>	---AlaAspGlySerProGlyLysAspGlyAlaArg	-----	752
Qy <td>418<td>GGACCGAGGAGGTCTACCCAGGACAGTACATTGCTTCCAGAGATCAAGTACAG</td><td> </td><td>477</td></td>	418 <td>GGACCGAGGAGGTCTACCCAGGACAGTACATTGCTTCCAGAGATCAAGTACAG</td> <td> </td> <td>477</td>	GGACCGAGGAGGTCTACCCAGGACAGTACATTGCTTCCAGAGATCAAGTACAG		477
Db <td>753<td>-----GlyLeuThrGlyPro</td><td>-----</td><td>757</td></td>	753 <td>-----GlyLeuThrGlyPro</td> <td>-----</td> <td>757</td>	-----GlyLeuThrGlyPro	-----	757
Qy <td>478<td>CTTCAACCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTCCCTGTAGCTCGT</td><td> </td><td>537</td></td>	478 <td>CTTCAACCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTCCCTGTAGCTCGT</td> <td> </td> <td>537</td>	CTTCAACCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTCCCTGTAGCTCGT		537
Db <td>758<td>IleGlyProGlyProAlaGly</td><td>-----AlaProGlyAspLysGlyGluAlaGlyPro</td><td>775</td></td>	758 <td>IleGlyProGlyProAlaGly</td> <td>-----AlaProGlyAspLysGlyGluAlaGlyPro</td> <td>775</td>	IleGlyProGlyProAlaGly	-----AlaProGlyAspLysGlyGluAlaGlyPro	775
Qy <td>538<td>ACATACCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGC</td><td> </td><td>597</td></td>	538 <td>ACATACCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGC</td> <td> </td> <td>597</td>	ACATACCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGC		597
Db <td>776<td>SerGlyPro</td><td>-----ProGly</td><td>786</td></td>	776 <td>SerGlyPro</td> <td>-----ProGly</td> <td>786</td>	SerGlyPro	-----ProGly	786
Qy <td>598<td>AGCCCAAGTGTAT</td><td>-----CTCACCCCTCTCACCCCTAAAGCTAT</td><td>636</td></td>	598 <td>AGCCCAAGTGTAT</td> <td>-----CTCACCCCTCTCACCCCTAAAGCTAT</td> <td>636</td>	AGCCCAAGTGTAT	-----CTCACCCCTCTCACCCCTAAAGCTAT	636
Db <td>787<td>AlaProGlyAspArgGlyGluAlaGlyProGlyProAlaGlyPheAlaGlyProPro</td><td> </td><td>806</td></td>	787 <td>AlaProGlyAspArgGlyGluAlaGlyProGlyProAlaGlyPheAlaGlyProPro</td> <td> </td> <td>806</td>	AlaProGlyAspArgGlyGluAlaGlyProGlyProAlaGlyPheAlaGlyProPro		806
Qy <td>637<td>-----CAACCTGTACAGTCTGTCTAAGCCACACCTGAGAGTAACCTTTTGGAT</td><td> </td><td>684</td></td>	637 <td>-----CAACCTGTACAGTCTGTCTAAGCCACACCTGAGAGTAACCTTTTGGAT</td> <td> </td> <td>684</td>	-----CAACCTGTACAGTCTGTCTAAGCCACACCTGAGAGTAACCTTTTGGAT		684
Db <td>807<td>GlyAlaAspGlyGlnProGly</td><td>-----AlaLysGlyGlu-ProGlyAspThrGlyValLysG</td><td>825</td></td>	807 <td>GlyAlaAspGlyGlnProGly</td> <td>-----AlaLysGlyGlu-ProGlyAspThrGlyValLysG</td> <td>825</td>	GlyAlaAspGlyGlnProGly	-----AlaLysGlyGlu-ProGlyAspThrGlyValLysG	825
Qy <td>685<td>GGGTTTGGCCCGCGCTGGACCACTAGCGGGCGGCTGGACGGCGATGAGCCCTGCC</td><td> </td><td>744</td></td>	685 <td>GGGTTTGGCCCGCGCTGGACCACTAGCGGGCGGCTGGACGGCGATGAGCCCTGCC</td> <td> </td> <td>744</td>	GGGTTTGGCCCGCGCTGGACCACTAGCGGGCGGCTGGACGGCGATGAGCCCTGCC		744

Db 825 yAspAlaGlyPro-ProGlyPro---AlaGlyProAlaGly-----p 838
QY 745 CTACCAACCCACATCTCAGACTGCGCACTTCCAGAGCCCTCCAGCAAGCACCCCTT 804
Db 838 roProGlyPro-----IleGlyAsnValGlyAlaProGlyProLysGlyProA 854
QY 805 AGGATACC-----AGGACCTCTTTCCCTTCGGCGGGGAAGTCATAGT 849
Db 854 rgGlyAlaAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyP 874
QY 850 AGGGAGCACCTCACACTTTCCTCAGGAGAGACACACAACCATGACCTCAGCCTGGGA 909
Db 874 roProGlyPro-----S 878
QY 910 CCCCAACTCCAGGCCCTCCAGCCCAACACTGCCAGCCAGCCCTGAAATGCAAGTCTT 969
Db 878 erGlyAsnAlaGlyProProGlyPro-ProGlyProValGlyLysGlyLysLys--- 896
QY 970 GTACGAGTTTGAAGCTAGGAACCCAGCGGAACTGACTGTGTGTC---GGGAGA----- 1020
Db 897 -----GlyProArgGly---GluThrGlyProAlaGlyArgProGly 909
QY 1021 GAAGCTGGAGGTTCTGGACACAGCAGCGGTGGTGGTGTGAAGATGAGCGGGAGC 1080
Db 910 GluValGlyProProGlyProProGlyPro-----AlaGlyGlu----- 922
QY 1081 GAGCGGCTACATTCCCAAGCAACATCTCGAGCCCT---ACAGCCGGGACCCCTGGGAC 1137
Db 923 -----LysGlySerProGlyAlaAspGlyProAlaGlySerProGlyThr 937
QY 1138 CCAGGCCAGTCACTCGGGTTCCAATGCTTCGACTTACGAGGCCCTGAAGAGT 1197
Db 938 ProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGlyLeuProGlyGlnArgGly 957
QY 1198 CACAGACTGGCTGCAGGACAGAACTTCTCCACTGCCCGGTGAGGACACTTGGTCCCT 1257
Db 958 GluArgGlyPheProGlyLeuPro-----GlyPro 967
QY 1258 GACGGGGAGCCAGCTACTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCCACAGGA 1317
Db 968 SerGlyGluProGlyLysGlnGlyProSer---GlySerSerGlyGluArgGlyProPro 986
QY 1318 GGCCCCACGAATCTCTCCCGCTGGAGCTGTCAGAAGGATGCTGG---GATAAGCCC 1374
Db 987 GlyProMetGlyProProGlyLeuAlaGlyProProGlyGluSerGlyArgGluGlySer 1006
QY 1375 TTAGGCACCAAGCTTAGACACTCCCAAGACCAAGCCCGCCCTGATGCAAGATGGCAGATCT 1434
Db 1007 ProGlyAlaGluGlySerProGlyArgAspGlyAlaProGlyAlaLysGlyAspArgGly 1026
QY 1435 GATACCCATTAGAGCCCGCAGAAATCTCTCTGATCCAGTTTGCAGCAAAACCCACA 1494
Db 1027 GluThrGlyProAlaGlyProProGlyAlaProGlyAlaProGlyAlaProGlyProVal 1046
QY 1495 CCCAGCTCACACAG-----CAAAACAATGGACAGGCCCGCCAGAGGCTGAAGCAACAGT 1548
Db 1047 GlyProAlaGlyLysAsnGlyAspArgGlyGluThrGlyProAlaGlyProAlaGlyPro 1066
QY 1549 GTC---CCTTCGGCTGTGTGA---GCCTCCCCA 1578
Db 1067 IleGlyProAlaGlyAlaArgGlyProAlaGlyPro 1078

Search completed: February 25, 2003, 19:41:10
Job time : 53.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 17:00:28 ; Search time 53 Seconds

(without alignments)

13295.870 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 3127

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Zgapop 6.0 , Zgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh

-Q=/cgn2_1/USPTO.spool/US09762021/runat_24022003_153038_7557/app.query.fasta.1.1863

-DB=SPTREMBL_21 -OFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL

-OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000

-USER=US09762021.ecgn.1.1.63@runat_24022003_153038_7557 -NCPU=6 -ICPU=3

-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120

-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7

-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phage:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rviro:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	%	Score	Match	Length	ID	Description
1	2416.5	77.3	594	4	Q96E47		Q96e47 homo sapien

2	2416	77.3	593	4	Q8TE67	Q8te67 homo sapien
3	2229	71.3	563	4	Q9H719	Q9h719 homo sapien
4	1606.5	51.4	600	11	Q91WL0	Q91wl0 mus musculu
5	597.5	19.1	729	11	Q9K30	Q9k30 mus musculu
6	588	18.8	743	4	Q9H6S3	Q9h6s3 homo sapien
7	584	18.7	715	4	Q9H8K9	Q9h8k9 homo sapien
8	574.5	18.4	596	4	Q9BBB7	Q9bbb7 homo sapien
9	572.5	18.3	723	4	Q8TE68	Q8te68 homo sapien
10	571.5	18.3	596	4	Q9NXH0	Q9nxh0 homo sapien
11	563	18.0	652	11	Q9D2M6	Q9d2m6 mus musculu
12	563	18.0	716	11	Q8R5F8	Q8r5f8 mus musculu
13	403	12.9	778	5	Q9VPD7	Q9vpd7 drosophila
14	402	12.9	778	5	Q9U3Z9	Q9u3z9 drosophila
15	399	12.8	437	4	Q9GZQ2	Q9gsq2 homo sapien
16	386	12.3	505	4	Q9BSQ2	Q9bsq2 homo sapien
17	356	11.4	929	5	O18249	O18249 caenorhabdi
18	352	11.3	732	5	O18250	O18250 caenorhabdi
19	351	11.2	712	5	Q95TJ6	Q95tj6 drosophila
20	341.5	10.9	714	5	Q9VEP7	Q9vep7 drosophila
21	333.5	10.7	472	11	Q8R0D6	Q8r0d6 mus musculu
22	329.5	10.5	327	4	Q9BUN9	Q9bum9 homo sapien
23	327.5	10.5	327	4	Q96K06	Q96k06 homo sapien
24	276.5	8.8	232	4	Q8WYW7	Q8wyw7 homo sapien
25	215	6.9	400	11	Q91VT7	Q91vt7 mus musculu
26	200.5	6.4	3084	12	O8U211	O8u211 pseudorabie
27	196.5	6.3	890	12	O8UZE1	O8uzel cercopithic
28	196.5	6.3	1548	4	Q9NYW9	Q9nyw9 homo sapien
29	196.5	6.3	2161	4	Q9Y566	Q9y566 homo sapien
30	196.5	6.3	5147	4	Q9Y6V0	Q9y6v0 homo sapien
31	195.5	6.3	3926	4	Q9UPA5	Q9upa5 homo sapien
32	195	6.2	645	4	Q96EX0	Q96ex0 homo sapien
33	195	6.2	1985	5	Q8T9N4	Q8t9n4 drosophila
34	193	6.2	606	11	Q8R0Y2	Q8r0y2 mus musculu
35	192.5	6.2	847	10	Q9XIB6	Q9xib6 arabidopsis
36	192	6.1	2167	11	Q9WV48	Q9wv48 rattus norv
37	191.5	6.1	437	12	O69146	O69146 human herpe
38	191	6.1	1300	12	O36421	O36421 alcelaphine
39	191	6.1	1963	5	Q9VSK5	Q9vsk5 drosophila
40	190.5	6.1	925	5	Q93238	Q93238 caenorhabdi
41	189.5	6.1	2158	11	Q9WU13	Q9wul3 rattus norv
42	189	6.0	1188	10	O41805	O41805 zea mays (m
43	189	6.0	2087	11	Q9W0E8	Q9wue8 rattus norv
44	188	6.0	1445	13	O93251	O93251 rana catesb
45	187	6.0	608	12	Q9Q5K9	Q9q5k9 herpesvirus

ALIGNMENTS

RESULT 1

Q96E47	Q96E47	PRELIMINARY;	PRT;	594 AA.
ID	Q96E47			
AC	Q96E47			
DT	01-DEC-2001	(TrEMBLrel. 19, Created)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)		
DE	Similar to hypothetical protein FLJ21522.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Srausberg R;			
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC012926; AAH12926.1; -			
DR	InterPro; IPR000050; PID_domain.			
DR	InterPro; IPR001452; SH3.			
DR	Pfam; PF00018; SH3; 1.			
DR	ProDom; PD000066; SH3; 1.			
DR	SMART; SM00462; PTB; 1.			
DR	PROSITE; PS00002; SH3; 1.			
SQ	SEQUENCE 594 AA; 66974 MW; F71E8F9B7564DEFE CRC64;			

Alignment Scores:
Pred. No.: 3,88e-184 Length: 594
Score: 2416.50 Matches: 458
Percent Similarity: 99.78% Conservativeness: 0
Best Local Similarity: 99.78% Mismatches: 0
Query Match: 77.28% Indels: 1
DB: 4 Gaps: 1

US-09-762-021A-1 (1-1710) x Q96E47 (1-594)

QY 2 GCAGAGCGACTGAAGACCAAGCTGCAAGAGCTCTGGAGAGAGAGCTGGAGCAA---AGA 58
DB 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluGluGluGlnSerArg 155
QY 59 CCTGAGCTTGGAGCTTCAGCCAGGCGCAGGACAGATGAGGGGCGCTCTATGGAAGG 118
DB 156 ProArgLeuGlyGlyLeuGlnProGlyGlnAspArgTrpArgGlyProAlaMetGluArg 175
QY 119 CCGCTCCCTATGAGCAGCAGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCAC 178
DB 176 ProLeuProMetGluGlnAlaArgTyrLeuGluProGlyLeuProGluGlnProHis 195
QY 179 CAGAGACCTTAGACACAGCTCCACCATCCCAAGGCCCTGCCAGCCACACCACT 238
DB 196 GlnArgThrLeuGluHisSerLeuProProSerProArgProArgHisThrSer 215
QY 239 GCCCGAGAACAAGTGCCTTTACTCTGCTCTCTCCAAAGCGCTCTTCTCCCGGAGGAC 298
DB 216 AlaArgGluProSerAlaPheThrLeuProProProArgSerSerProGluAsp 235
QY 299 CCAGAGAGGAGCAGGAGTGTCTGAACCATGTCTTAAGGACATTTGAGTGTTCATGGGA 358
DB 236 ProGluArgAspGluGluValLeuAsnHisValLeuArgAspGluLeuPheMetGly 255
QY 359 AGCTGGAGAAGCCAGCAAGACCCAGCAGAGAGAAATTTGGGAAAAAACAAG 418
DB 256 LysLeuGluLysAlaGlnAlaLysThrSerArgLysLysPheGlyLysLysAsnLys 275
QY 419 GACCAGGAGGTCTCACCCAGCAGACATGTACTGCTCTCCAGAGATCAAGTACAGC 478
DB 276 AspGlnGlyLeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysThrSer 295
QY 479 TTCAACCTCTCTGGAGGCTGCCACCTGGCTGAAGCAGACAGTGCCTCGAGTCTGTA 538
DB 296 PheAsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuVal 315
QY 539 CACATCTCTCAAGTCCCTGAATTCATCTCTGGCCAGGTGCCCTGAGGCTGGCTAGCA 598
DB 316 HisIleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAla 335
QY 599 GCCCAAGTATCTCACCCCTCTCACCCCTAAAGCTATCAAGCTGCTACAGTCTGTCTA 658
DB 336 AlaGlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeu 355
QY 659 AGCCCACTCTGAGAGTAACTTTGGATGGGCTGGGCCAGCCCTGAGCAGCCGGGC 718
DB 356 SerProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrSerArgAla 375
QY 719 GACTGGACAGGCGATGAGCCCTGCCCTACCAACCCACATTTCTCAGATGACGTGCAACTT 778
DB 376 AspTrpThrGlyAspGluProLeuProTyrGlnProThrPheSerAspAspTrpGlnLeu 395
QY 779 CCAGAGCCCTCCAGCAACCCCTTAGCATACAGGACCTGTTCCTCTCCGGCGGGA 838
DB 396 ProGluProSerSerGlnAlaProLeuGlyTyrGlnAspProValSerLeuArgGly 415
QY 839 ACTCATAGTTAGGAGCAGCTCACACTTTCTCTCAGGAGAAGACACACCAACCATGACCTT 898
DB 416 SerHisArgLeuGlySerThrSerHisPheProGlnGluLysThrHisAsnHisAspPro 435
QY 899 CAGCCTGGGAGCCCAACTCCAGGCGCTCCAGCCCAACCTGCGCCAGCCACCTCGAAA 958

DB 436 GlnProGlyAspProAsnSerArgProSerSerProLysProAlaGlnProAlaLeuLys 455
QY 959 ATGCAAGTCTTCTACGAGTTTCAAGCTAGGAACCCACGGAACCTGACTGCTGCTCCAGGA 1018
DB 456 MetGlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThrValGlnGly 475
QY 1019 GAGAAGCTGGAGTCTTGACACACAGCGGTGGTGGTGAAGAAATGAGCGGGA 1078
DB 476 GluLysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGly 495
QY 1079 CGGAGCGCTACATTCACAGCAACATCTGGAGCCCTACACCCGGGGACCTCGGACC 1138
DB 496 ArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThr 515
QY 1139 CAGGCGCAGTCAACCTCTCGGGTTCACATCTTCACCTTAGCTGAGGCTCAAGAGGTC 1198
DB 516 GlnGlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluVal 535
QY 1199 ACAGACTGGCTCAGCAGCAGAACTTCTCCACTGCCAGCGTGAGGACACTTGGGTCCTCG 1258
DB 536 ThrAspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeu 555
QY 1259 AGCGGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGGAG 1318
DB 556 ThrGlySerGlnLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGlu 575
QY 1319 GCCCAGCAATCTCTCCCGGCTGGAGGCTGTCAAGAGATGCTGGGATAAGCCCT 1375
DB 576 AlaProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 594

RESULT 2

Q8TE67 PRELIMINARY: PRT: 593 AA.
AC Q8TE67: 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor pathway substrate 8 related protein
DE 3.
GN EPS8R3
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074930; AAL76119.1; -
KW Receptor.
SQ SEQUENCE 593 AA; 66891 MW; E33134671B9150AF CRC64;

Alignment Scores:
Pred. No.: 4,25e-184 Length: 593
Score: 2416.00 Matches: 456
Percent Similarity: 99.78% Conservativeness: 1
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 77.26% Indels: 0
DB: 4 Gaps: 0

US-09-762-021A-1 (1-1710) x Q8TE67 (1-593)

QY 2 GCAGAGCGACTGAAGACCAAGCTGCAAGAGCTCTGGAGAGAGAGCTGGAGCAAACCT 61
DB 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluGluGlnArgPro 155
QY 62 CCAGCTGGAGGCTTCAGCCAGGCGCAGACATGAGGGGCGCTGCTATGCAAGGCGC 121
DB 156 ArgLeuGlyGlyLeuGlnProSerGlnAspArgTrpArgGlyProAlaMetGluArgPro 175


```
QY 122 CTCCTATGAGCAGCGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCGACAG 181
Db 176 LeuProMetGluGlnAlaArgTyrLeuGluProGlyIleProProGluGlnProHisGln 195
QY 182 AGGACCTAGACGACGCTCCACCATCCCAAGCGCCCTCCAGCGCCACACAGTGC 241
Db 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
QY 242 CCAGAACCAAGTCTTACTCTGCTCTCCCAAGCGGCTCTCTTCCCGGAGGACCCA 301
Db 216 ArgGluProSerAlaPheThrLeuProProProArgArgSerSerProGluAspPro 235
QY 302 GAGAGGCGAGGAAGTCTGAACCATGCTCTTAAGGAGCATTGAGCTGTTCATGGGAAG 361
Db 236 GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
QY 362 CTGGAGAGGCCCGCAGCAAGCAGCAGGAGAGAAATTTGGGAAAAAACAAGGAC 421
Db 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysLysPheGlyLysAsnLysAsp 275
QY 422 CAGGAGGTCTCACCGCAGCAGTACATTGACTCTCCAGAACATCAAGTACAGCTTC 481
Db 276 GlnGlyGlyLeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysHisSerPhe 295
QY 482 AACCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCTCGTACAC 541
Db 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
QY 542 ATCCTCTTCAAGTCTCAACTTCATCTGCGCCAGGTCGCCCTGAGCGTGGCTACGACG 601
Db 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAlaAla 335
QY 602 CAAGTGATCTCACCCCTCTCAACCTTAAAGCTATCAACCTGCTACAGTCTCTTAAGC 661
Db 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeuSer 355
QY 662 CCACCTGAGAGTAACCTTTGGATGGGTGGGCCAGCGCTGGACACTAGCGGCGCAG 721
Db 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrThrSerArgAlaAsp 375
QY 722 TGGACAGGCGATGACCCCTGCCCTACCAACCCCATCTCTCAGATGACTGGCACTTCCA 781
Db 376 TrpThrGlyAspGluProLeuProTyrGlnProThrPheSerAspTrpGlnLeuPro 395
QY 782 GAGCCCTCGAGCAAGCACCTTAGGATACAGGACCGCTTTTCCCTTCGGCGGGAAGT 841
Db 396 GluProSerSerGlnAlaProLeuGlyTyrGlnAspProValSerLeuArgArgGlySer 415
QY 842 CATAGGTTAGGAGCACCTCACACTTCTCTCAGGAGAGACACACACCATGACCCCTCAG 901
Db 416 HisArgLeuGlySerThrSerHisPheProGlnGluLysThrHisAsnHisAspProGln 435
QY 902 CTTGGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATG 961
Db 436 ProGlyAspProAsnSerArgProSerProLysProAlaGlnProAlaLeuLysMet 455
QY 962 CAAGTCTTGTACAGTTTGAAGCTAGGAACCCACGGGAACCTGACTGTGTCTCAGGAGAG 1021
Db 456 GlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGlu 475
QY 1022 AAGCTGGAGGTTCCTGGACACAGCAAGCGGTGGTGGCTGGTGAAGAATCAGCGCGGAGG 1081
Db 476 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 495
QY 1082 ACGGCTACATTCCAGCAACATCTCTGGAGCCCTTACAGCCGGGAGCCCTGGGACCCAG 1141
Db 496 SerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 515
QY 1142 GCCAGTCAACCTCTCTGGTTCCTAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCA 1201
Db 516 GlyGlnSerProSerArgValPrometLeuArgLeuSerSerArgProGluValThr 535
QY 1202 GACTGGCTGCAGGCGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGGTCCCTGACG 1261
```

```
Db 536 AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 555
QY 1262 GGGACCCAGCTACTTCGCATAGACCTGGGAGCTTACAGATGCTATGTCCAGAGGCC 1321
Db 556 GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAla 575
QY 1322 CCACCAATCTGTCTCCGCTGGAGGCTGTCTCAGAAGGATGCTGGGATAAGCCCT 1375
Db 576 ProArgIleLeuSerArgLeuGluAlaValArgMetLeuGlyIleSerPro 593
RESULT 3
Q9H719
ID Q9H719 PRELIMINARY; PRT; 563 AA.
AC Q9H719;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE CDNA: FLJ21522 fis, clone COL05884.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK0251175; BAB15081.1; -.
DR HSSP; Q08509; 1AQJ.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 563 AA; 63527 MW; A26915ED19057E9B CRC64;
Alignment Scores:
Pred. No.: 3.36e-169 Length: 563
Score: 2229.00 Matches: 428
Percent Similarity: 93.45% Conservative: 0
Best Local Similarity: 93.45% Mismatches: 0
Query Match: 71.28% Indels: 30
DB: 4 Gaps: 1
US-09-762-021a-1 (1-1710) x Q9H719 (1-563)
QY 2 GCAGAGCGACTCAAGCAGCAGCTCGCAGAGGCTCGGAGGAAGCTGGAGCAAGACCT 61
Db 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluLeuGluGlnArgPro 155
QY 62 CGACTTGGAGGCTTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGCTATGTAAGAGCGG 121
Db 156 ArgLeuGlyLeuGlnProGlyGlnAspArgTrpArgGlyProAlaMetGluArgPro 175
QY 122 CTCCTATGAGCAGCAGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCGACAG 181
Db 176 LeuProMetGluGlnAlaArgTyrLeuGluProGlyIleProProGluGlnProHisGln 195
QY 182 AGGACCTAGACGACGCTCCACCATCCCAAGCGGCTCTCTTCCCGGAGGACCCA 241
Db 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
QY 242 CCAGAACCAAGTCTTACTCTGCTCTCCCAAGCGGCTCTCTTCCCGGAGGACCCA 301
```

```
Db 216 ArgGluProSerAlaPheThrLeuProProArgArgSerSerProGluAspPro 235
QY 302 GAGAGGACGAGAGTGTGTAACCATGCTCTAAAGGACATTGAGCTGTTCATGGAAAG 361
Db 236 GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
QY 362 CTGGAGAAGCCCGAGCAAGACACGACGAGAGAAATTTGGGAAAAAACAAGGAC 421
Db 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysLysPheGlyLysLysAsnLysAsp 275
QY 422 CAGGGAGGTCTCACCAGCAGCAGTACATTGCTGCTCCAGAAAGATCAAGTACAGCTTC 481
Db 276 GlnGlyGluThrGlnAlaGlnIleAspCysPheGlnLysIleLysIleLysIleLysIle 295
QY 482 AACCTCTGGGAGGCTGCCACCTGGCTGAAGGACAAAGTGCCTCAGCTCGTACAC 541
Db 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
QY 542 ATCTCTTCAAGTCCCTGAACCTTCATCCTGGCCAGGTGCCCTGAGGCTGGCCCTAGCAGCC 601
Db 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAlaAla 335
QY 602 CAACTGATCTACCCCTCTCACCCTTAAGCTATCAAGCTATCAAGCTGCTACAGTCTGTAAAC 661
Db 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeuSer 355
QY 662 CCACCTGAGAGTAACCTTTGGATGGGTTGGGCCAGCTGGACCACTAGCCGGGCGGAC 721
Db 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaThrTrpThrSerArgAlaAsp 375
QY 722 TGGACAGGCGATGAGCCCTGCCCTACCAACCCACATTCCTCAGATGACTGGCAACTTCCA 781
Db 376 TrpThrGlyAspGluProLeuProTrpGlnProThrPheSerAspAspTrpGlnLeuPro 395
QY 782 GAGCCCTCAGCAACGACCCCTTAGATACAGGACCCCTGTTCCCTTCGGGGGGAAGT 841
Db 396 GluProSerSerGlnAlaProLeuGlyTrpGlnAspProValSerLeu----- 411
QY 842 CATAGTTAGGGAGCACCTCACATTTTCTCAGGAGAAGACACACACCAACCATGACCTCAG 901
Db 411 ----- 411
QY 902 CTGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATG 961
Db 412 -----ArgProSerSerProLysProAlaGlnProAlaLeuLysMet 425
QY 962 CAACTCTTGTACAGTTTGAAGCTAGGAACCCACGCGGAAGTACTGTGTCTCAGGAGAG 1021
Db 426 GlnValLeuTyrrGluPheGluAlaArgAsnProArgGluLeuThrValGlnGlyGlu 445
QY 1022 AAGCTGGAGGTCTGGACCAACAGCAGCGTGTGGCTGGTGAAGAAATCAGCGGGGACGG 1081
Db 446 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 465
QY 1082 AGCGGTACATTCCAAGCAACATCTCTGGAGCCCTACAGCCGGGGACCCCTCGGAGCCAG 1141
Db 466 SerGlyTyrrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 485
QY 1142 GCCAGTACCCCTCTCGGGTTCCAAATGCTTCGACTTAGCTAGCTCGAGGCTCAAGAGGTCA 1201
Db 486 GlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluValThr 505
QY 1202 GACTGGCTCAGGACAGAACTTCTCCACTGGCCAGCGGTGAGACACTTGGGTCCTGAGC 1261
Db 506 AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 525
QY 1262 GGGAGCCAGCTACTTCGCATAGACCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCC 1321
Db 526 GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAla 545
QY 1322 CCACGAATCTCTCCCGGCTGAGGCTGTCCAGAAGATCTGGGGTAAGCCCT 1375
Db 546 ProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 563
```

RESULT 4

```
Q91WL0
ID Q91WL0 PRELIMINARY; PRT; 600 AA.
AC Q91WL0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ21522 (Epidermal growth factor
DE receptor pathway substrate 8 related protein 3).
GN AI504489 OR EPS8R3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014734; AAL14734.1; -.
DR MGD; MGI:2139743; AI504489.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Hypothetical protein; Receptor.
SQ SEQUENCE 600 AA; 68215 MW; 4AD0F34FCCCCDD39 CRC64;

Alignment Scores:
Pred. No.: 1,38e-119 Length: 600
Score: 1606.50 Matches: 318
Percent Similarity: 78.97% Conservative: 50
Best Local Similarity: 68.24% Mismatches: 87
Query Match: 51.38% Indels: 11
DB: 11 Gaps: 6

US-09-762-021a-1 (1-1710) x Q91WL0 (1-600)
QY 2 GCAGAGCGACTGAAGACCAAGCTCTGGAGAGAGAGCTGGAGCAAGAGCTT 61
Db 136 AlaGluGlnLeuArgThrSerLeuGlnLysAlaLeuGluGluGluGluArgPro 155
QY 62 CCACCTGGAGGCTTCAGCCAGCCAGGACACATGAGGGGGCTGTATGGNAAGCGCG 121
Db 156 ArgPheGlyValIleHisProSerGlnAspArgTrpLysGlyProProLeuGluArgPro 175
QY 122 CTCCTCTATGAGCAGCAGCTATCTGGAGCGC-----GGGATCCCT 163
Db 176 LeuProIleGlnAlaProProLeuGluGlnArgPheSerProGluHisArgPhePro 195
QY 164 CCAGAACAGCCCAAGAGGACCCCTAGACAGACGCTCCCAACCATCCCAAGCCCTG 223
Db 196 ProGluGlnProHisAsnMetThrSerGluArgSerIleSerProSerArgSerLeu 215
QY 224 CCAGCCCAAGAGGCGGAGAACCAAGTGCCTTTACTCTG---CCTCTCCAAGCGCG 280
Db 216 ThrHisTyrrProSerAlaArgGluProAsnGlyPheThrLeuProProProArgArg 235
QY 281 TCCTCTTCCCGGAGGACCCAGAGGAGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGAC 340
Db 236 AlaProSerProGluAspProGluArgAspGluGluValLeuAsnHisValLeuArgAsp 255
QY 341 ATTAGCTCTTTCATGGGAAAGCTGGAGAGAGGCCCGCCAGCAAGACCAAGAGAG-- 397
```

```
Db 256 lileGluLeuPheAlaGlyLysLeuLysGluValGlnAlaArgAsnSerHisLysLysThr 275
Qy 398 MAATTGGG---AAAAAACAACAGCAGGAGGTCTCACCCAGGACAGTACATTGAC 454
Db 276 LysLeuGlyArgLysLysSerLysAsnGlyIleThrGlnAlaGlyTrpIleAsp 295
Qy 455 TGTTCACAGAGATCAAGTACAGCTTCAACCTCTCGGAAGGCTGCCACCTGGCTGAAG 514
Db 296 CysPheGlnLysIleLysLysSerPheAsnLeuLeuGlyLysLeuAlaLeuArgMetGln 315
Qy 515 GAGACAAGTCCCTGAGCTCGTACACATCTCTCAAGTCCCTCAAGTCACTCATCTGCC 574
Db 316 GluThrSerAlaProGluPheValGlyLeuIlePheGlnThrLeuLysPheIleLeuSer 335
Qy 575 AGGTCCCTGAGGCTGAGCAGCCCAAGTGATCTCACCCCTCTCACCCCTCAAAAGCT 634
Db 336 GlnCysProGluAlaGlyLeuProAlaLysValIleSerProLeuLeuThrProLysAla 355
Qy 635 ATCAACCTGCTACAGTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGTGGGC 694
Db 356 IleAspLeuLeuGlnSerCysLeuSerProProGluAspThrLeuTrpLysSerLeuGly 375
Qy 695 CCAGCTGGACCACTAGCGGGCCGACTGGACGCGGATGAGCCCTGCCCTACCAACC 754
Db 376 ThrSerTrpThrThrSerTrpAlaAspTrpThrGlySerGluProProProTrpGlnPro 395
Qy 755 ACATTCTCAGATGATGCTGCACTTCCAGAGCCCTCCAGCCAGCACCTTAGGATACAC 814
Db 396 ThrPheTrpAspGlyTrpGlnIleProGlnProArgSerMetMetProIleThrAsnGln 415
Qy 815 GACCCTGTTTCCCTCGCGGGGAAGTCATAGTTAGGAGCACCTCACACTTTCTCTCAG 874
Db 416 AspSerIleSerLeu---ArgGlySer---ArgMetArgSerSerLeuHisPheProArg 433
Qy 875 GAGAAGACACACACCACTGACCTCAGCTGCTGGGAGCCCACTCCAGCCCTCCAGCCCC 934
Db 434 AspGluProTrpAsnHisAsnProGluTrpGluAspSerAsnLeuProLeuSerSerPro 453
Qy 935 AAACCTGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCA 994
Db 454 SerProGlyArgAlaAlaLeuLysMetGlnValLeuTrpGluPheGluAlaArgAsnAla 473
Qy 995 CGGGAACCTGCTGGTCCAGGAGAGAGCTGGAGGTTCTGGACACAGCAAGCGTGG 1054
Db 474 GlnGluLeuThrValAlaGlnGlyIleLeuGluValLeuAspGlnSerLysArgTrp 493
Qy 1055 TGGCTGGTCAGATGAGCGGGGAGCGGAGCGGCTACATTCCAAGCAACATCTGGAGCC 1114
Db 494 TrpLeuValLysAsnGluAlaGlyLeuThrGlyTrpIleProSerAsnIleLeuGluPro 513
Qy 1115 CTACAGCCGGGACCCCTGGGAGCCAGCCAGTCACCTCTCGGTTTCAATGCTCGA 1174
Db 514 LeuProAlaGlyAlaProArgGlyHisArgGlnProSerPheArgAlaProMetLeuArg 533
Qy 1175 CTAGCTCGAGCCCTGAAGAGTTCACAGCTGGCTGCGAGGAGAGAACTTCTCCACTGCC 1234
Db 534 LeuSerSerLysProGluGluValThrAlaTrpLeuGlnAlaGluAsnPheSerThrVal 553
Qy 1235 ACGGTGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATTAAGACCTGGGAG 1294
Db 554 ThrValArgThrLeuGlySerLeuMetGlySerGlnLeuLeuHisMetArgProGlyGlu 573
Qy 1295 CTACAGATGCTATGCCACAGGAGCCCAAGTCTCTGTCGCCGCTGAGGCTCTCAGA 1354
Db 574 LeuGlnMetLeuLysProGlnGluAlaProArgIleGlnAlaArgLeuAspAlaValArg 593
Qy 1355 AGGATGCTGGGATAAGC 1372
Db 594 ArgMetLeuGlyMetThr 599
RESULT 5
Q99K30
```

```
ID AC Q99K30 PRELIMINARY; PRT; 729 AA.
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ21935.
GN A1042819.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; BC005492; AA005492.1; -.
DR HSP; Q08509; IAOJ.
DR MGI; MGI:2138828; A1042819.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS00002; SH3; 1.
DR Hypothetical protein; SH3 domain.
KW Hypothetical protein; SH3 domain.
SQ SEQUENCE 729 AA; 82229 MW; A11DE82FF2C0BD18 CRC64;

Alignment Scores:
Pred. No.: 3,7e-39 Length: 729
Score: 597.50 Matches: 173
Percent Similarity: 46.21% Conservative: 95
Best Local Similarity: 29.83% Mismatches: 189
Query Match: 19.11% Indels: 124
DB: 22

US-09-762-021A-1 (1-1710) x Q99K30 (1-729)
Qy 5 GAGCGACTGAGACAGCAGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGAGCTCGA 64
Db 178 LysLysMetArgProGlnThrLeuLysGlyHisGlnGluLysIleArgGlnArgGlnSer 197
Qy 65 CTGGAGGCTTCAGCGCCAGGACAGATGGAGGGGCGCTGCTATGAAAGGCGCTC 124
Db 198 Ile-----LeuProProGln-----SerProAla-----ProIle 208
Qy 125 CCTATGGAGCAGCAGCTATCTGGAGCGGGGATCTCCAGAACAGCCCCACAGAGG 184
Db 209 ProPheGlnArg-----GlnProGlyAspSerProGlnAlaLysAsnArgVal 224
Qy 185 ACCCTAGAGCAGCCTCCACCATCCCAAGGCCCTGCCACGCCACACAGTCCCGCA 244
Db 225 GlyLeu-----ProLeuProValPro----- 231
Qy 245 GAACCAAGTGCTTTACTCTGCTCTCT--CCAAGGCGGTCTCTTCCCGGAGGCCCA 301
Db 232 -----PheSerGluProGlyTrpArgArgGluSerGlnAspGluGluPro 247
Qy 302 -----GAGAGGACGAGGAAGTGTGCAACCATGTCTCTAAGG 337
Db 248 ArgAlaValLeuAlaGlnArgIleGluLysGluThrGlnIleLeuAsnCysThrLeuAsp 267
Qy 338 GACATTGAGCTTTCATGGGAAAGCTGGAGAGGCCCGCAG-----GCAAGACACAGC 388
Db 268 AspIleGluTrpPheValAlaArgLeuGlnLysAlaAlaGluAlaPheLysGlnLeuAsn 287
Qy 389 AGGAAGAAGAAATTTGGGAAAAAACAAGAC-----CAGGAGGCTCTACC--- 436
Db 288 GlnArgLysLysGlyLysLysAsnLysLysGlyProAlaGluGlyValLeuThrLeu 307
Qy 437 -----CAGGCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGC 478
```

```
Db 308 ArgAlaArgProSerGluGlyGluPheValAspCysPheGlnLysThrLysLeuAla 327
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 479 TTCAACCTCTGGAGGCTGCCACCTGGCTCAAGGACAGTCCCTGAGTCGTA 538
      ||||| ||||| ::::: ::::: ||||| ||||| ||||| |||||
Db 328 IleAsnLeuAlaLysLeuGlnLysHisIleGlnAsnProSerAlaAlaGluLeuVal 347
      ||||| ||||| ::::: ::::: ||||| ||||| ||||| |||||
QY 539 CACATCCTCTCAAGTCCCTGAATTCATCTCGGCGGAGTGGCCTGAGGCTGGCCTAGCA 598
      ||||| ||||| ::::: ::::: ||||| ||||| ||||| |||||
Db 348 HisPheLeuPheGlyProLeuAspLeuIleIleAsnThrCysGlySerProAspIleAla 367
      ||||| ||||| ::::: ::::: ||||| ||||| ||||| |||||
QY 599 GCCCAAGTATCATCCCTCCCTCAACCTTAAAGCTATCAACCTGTACAGTCTGTCTA 658
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 368 ArgSerValSerSerProLeuLeuSerThrAspAlaValSerPheLeuArgGlyHisLeu 387
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 659 AGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCGGAGCTGGACCACTAGCCGGCC 718
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 388 ValProLysGluMetThrLeuThrGluSerLeuGlyGluThrTrpMetArgProArgSer 407
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 719 GACTGGACAGGCGATGAGCCCTGGCC---TACCAACCCACATTCACAGATGACTGGCAA 775
      ::::: ::::: ||||| ||||| ||||| ||||| ||||| |||||
Db 408 GluTrpProArgGluProGlnValProLeuTyrrValProLysPheArgSerGlyTrpGlu 427
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 776 CTTCCA---GAGCCCTCCAGCCAGCACCCCTTAGGATACAGGAC-----CCT 820
      ||||| ::::: ||||| ||||| ||||| ||||| ||||| |||||
Db 428 ProProLeuAspValLeuGlnGluAlaProTrpGluValGluGlyLeuAlaSerValPro 447
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 821 GTTTCCTTCGGCGGGAAGTCACTAGTTAGGAGCACCTCACACTTCTCAGGAGAAG 880
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 448 SerAspGlnLeuThrProLysAsnArgLeu---SerValArgHisSerProLysHisSer 466
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 881 ACACAACCAACCTGACCTCAGCCTGGGGAC-----CCCAACTCCAGGCCCTCCAGC 931
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 467 LeuSerSerGluSerGlnAlaProGluAspIleAlaProGlySerSerProHisAla 486
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 932 CCCAAACCTGCCAGCCAGCCCTGAAATG-----CAAGCTGTACAGGTTT 979
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 AsnArgGlyTyrrGlnProThrProAlaMetThrLysTyrrValLysIleLeuTyrrAspPhe 506
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 980 GAAGCTAGGAAGCCCGGGAAGTCACTGTGGTCCAGGAGAGAGCTGGAGTCTGGAC 1039
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 507 ThrAlaArgAsnAlaAsnGluLeuSerValLeuLysAspGluValLeuValLeuGlu 526
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1040 CACAGCAAGCGTGGTGGTGAAGATGAGCGGGGAGCGGCTACATTCACAAGC 1099
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 527 AspGlyArgGlnTrpLysLeuArgAsnArgSerGlyGlnAlaGlyTyrrValProCys 546
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1100 AACATCTG-----GAGCCCTACAG----- 1120
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 AsnIleLeuAlaGluAlaArgGlnGluAspValGlyAlaProLeuGluGlnSerGlyGln 566
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1121 -----CCGGGGACCCCT----- 1132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 567 LysTyrrTrpGlyProAlaSerProThrHisLysLeuProIlePheAlaGlyAsnLys 586
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1132 ----- 1132
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 587 GluGluLeuIleHisMetAspGluValAsnAspGluLeuMetLysLysIleSerHis 606
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1133 ---GGGACCCAGCCAGTCACCTCTCGGGTT-----CCAATG----- 1168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 IleLysThrGlnProGlnArgAsnPheArgValGluArgSerGlnProValHisLeuPro 626
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1169 CTTTCGACTTAGCTCAGGCTGAAGAGTCAACAGCTGGCTGCAGGCAGAGAACTTCTCC 1228
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 627 LeuThrPheGluSerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPheSer 646
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1229 ACTGCCAGGTGAGACACTTGGGTCCCTCAGCGGGAGCCAGCTACTTCGCATAAGACCT 1288
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 647 AlaArgIleValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLys 666
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1289 GGGAGCTACAGATGTATGTCACAGGAGGCCCCCAGTAATCTGTCCCGCTGGAGGCT 1348
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 667 GluGluLeuLysLysValCysGlyGluGlySerArgValTyrrSerGlnLeuThrVal 686
QY 1349 GTCAGAAGGATCCTGGG-CATAAGCCCTTAGCCACAGCTTAGACACCTCCAAGAACCAG 1407
      ::::: ||||| ::::: ||||| ||||| ||||| ||||| |||||
Db 687 GlnLysAlaPheLeuGluLysGlnSerGlySerGluLeuGluLysLeuMetSerLys 706
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1408 GCCCCGCTGATCAAGATGGCAGATCTGATACCCATTAGAGCCCGGAGAAATTCCTTCT 1467
      ::::: ||||| ::::: ||||| ||||| ||||| |||||
Db 707 IleArgArgAlaGluAspSerTyrrThrSerGlnHisThrSerProGluSerGluGlyAla 726
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
Q9H6S3
ID Q9H6S3 PRELIMINARY; PRT; 743 AA.
AC Q9H6S3;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CDNA: FLJ21935 fis, clone HEP04373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK025588; BAB15180.1; -.
DR HSSP; Q08509; 1A07.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS50002; SH3; 1.
DR SH3 domain.
KW SH3 domain.
SQ SEQUENCE 743 AA; 83792 MW; DB01D8F6363A1F80 CRC64;
```

Alignment Scores:

Pred. No.:	2,12e-38	Length:	743
Score:	588.00	Matches:	162
Percent Similarity:	44.72%	Conservative:	71
Best Local Similarity:	31.09%	Mismatches:	136
Query Match:	18.80%	Indels:	132
DB:	4	Gaps:	18

US-09-762-021A-1 (1-1710) x Q9H6S3 (1-743)

```
QY 131 GAGCAGGACGGTATCTGGAGCGGGGATCCCTCCAGAACAGCCGCCACGAGAGCCCTA 190
      ||||| ||||| ::::: |||||
Db 218 GluLysIleArgGlnArgGlnSerPheLeuProPro----- 229
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 GAGCAGAGCCTCCCAACATCCCAAGCCCTCCCA---CGCCACACCACTGCCCCAGAA 247
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 230 -----ProGlnGlyProAlaProIleProPheGlnHisArgGlyGlyAspSer 245
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 248 CCAAGTGCC-----TTTACTCTGCCTCCTCCAAGCGCG 280
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 246 ProGluAlaLysAsnArgValGlyProGlnValProLeuSerGluProGlyPheArg 265
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 281 TCCTCTTCCCGCAGACCCA-----GAGAGGGACGAGGAA 316
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 266 ArgGluSerGlnGluProArgAlaValLeuAlaGlnLysIleGluLysGluThrGln 285
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 317 GTCTCAACCATCTCCCTAAGGACATTGAGCTGTTCTTCAAGGAAAGCTGGAGAGGCC--- 373
      ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 286 IleLeuAsnCysAlaLeuAspAspIleGluTrpPheValAlaArgLeuGlnLysAlaAla 305
```

```
QY 374 -----CAGCAAGACAGCAGGAGAGAAATTTGGGAAAAA 412
Db 306 GluAlaPheLysGlnLeuAsnGlnArgLysLysGlyLysLys-----GlyLysLys 323
QY 413 ACAAGGACAGGAGGAGTCTAC-----CAGCACACATACATT 451
Db 324 AlaProAlaGluGlyValLeuThrLeuArgAlaArgProProSerGluGlyGluPheLe 343
QY 452 GACTGCTCCAGAGATCAAGTACAGTCAACCTCTCGGAGAGCTGCCACCTGGCTG 511
Db 344 AspCysPheGlnLysLeuAlaLeuAsnLeuLeuAlaLysLeuGlnLysHisLe 363
QY 512 AAGGAGACAAGTGCCTGAGCTGCTGACACATCTCTTCAAGTCCCTGAACATCATCTG 571
Db 364 GlnAsnProSerAlaAlaGluLeuValHisPheLeuGlyProLeuAspLeuLeVal 383
QY 572 GCCAGGTGCCGTGAGCTGCTAGCAGCCCAAGTGTCTCACCCCTCTCACCCCTAAA 631
Db 384 AsnThrCysSerGlyProAspLeuAlaArgSerValSerCysProLeuLeuSerArgAsp 403
QY 632 GCTATCAACCTGCTACAGTCCCTGTCTAAGCCCACTGAGAGTAACTTTGGATGGGTG 691
Db 404 AlaValAspPheLeuArgGlyHisLeuValProLysGluMetSerLeuTrpGluSerLeu 423
QY 692 GCCCAGCTGGACACTAGCCGCGGAGTGGACAGCGGATGAGCCCTGCC-----TAC 748
Db 424 GlyGluSerTrpMetArgProArgSerGluTrpProArgGluProGlnValProLeuTyr 443
QY 749 CAACCCACATCTCAGATGACTGCACTTCCA-----781
Db 444 ValProLysPheHisSerGlyTrpGluProProValAspValLeuGlnGluAlaProTrp 463
QY 782 -----GAGCCCTCAGCAGACACCTTAGGATACAGGACCCGTTTCCCTCGCGG 835
Db 464 GluValGluGlyLeuAlaSerAlaProIleGluGluValSerProValSer-----Arg 481
QY 836 GGAAGTCATAGTTAGGAGGACCTCACACTTTCTCAGGAGAGACACACAACCATGAC 895
Db 482 GlnSerIleArg---AsnSerGlnLysHisSerPro-----ThrSerGluProThr 497
QY 896 CTTACGCTGGGAC-----CCCACTCCAGGCTCAGCCCAACCTGCCAG 946
Db 498 Pro---ProGlyAspAlaLeuProProValSerProHisThrHisArgGlyTyrGln 516
QY 947 CCA-----GCCTGAAATGCAAGTCTGTACGAGTTTGAGCTAGGAGAACCA 994
Db 517 ProThrProAlaMetAlaLysTyrValLysIleLeuTyrAspPheThrAlaArgAsnAla 536
QY 995 CGGGAAGTACTGTGCTCAGGAGAGAGTGGAGGTTCTGACACAGCAAGCGGTGG 1054
Db 537 AsnGluLeuSerValLeuLysAspGluValLeuGluValLeuGluAspGlyArgGlnTrp 556
QY 1055 TGGCTGGTGAAGATGAGCGGAGCGGCTACATTCACATCCAGCAACATCTG-----1108
Db 557 TrpLysLeuArgSerArgSerGlyGlnAlaGlyTyrValProCysAsnIleLeuGlyGlu 576
QY 1108 -----1108
Db 577 AlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGlyLysTyrTrpGlyPro 596
QY 1109 -----GAGCCCTCAGCAGCGGAGCCCTGGAGCCAGCGGCAG-----1147
Db 597 AlaSerProThrHisLysLeuProProSerPheProGlyAsnLysAspGluLeuMetGln 616
QY 1147 -----1147
Db 617 ArgMetAspGluValAsnAspGluLeuIleArgLysIleSerAsnIleArgAlaGlnPro 636
QY 1148 -----TCACCTCTCGGTTCCAATCTTCGACTTAGC 1180
Db 637 GlnArgHisPheArgValGluArgSerGlnProAlaSerGlnPro---LeuThrTyrGlu 655
QY 1181 TCGAGGCTGAAGAGTTCACAGACTGGCTGACAGGAGAGAACTTCTCCACTGCCACGGTG 1240
```

```
Db 656 SerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPheSerProArgIleVal 675
QY 1241 AGGACACTGGCTCCTCAGCGGGAGCCAGCTACTTCGCATAGAGCTGGGAGCTACAG 1300
Db 676 GluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLysGluGluLeuLys 695
QY 1301 ATGCTATCTCCACAGAGAGCCCAAGTCTCTGTCCTCCGCTGGAGCTGTCTCAGAAGGATG 1360
Db 696 LysValCysGlyGluGluGlyValArgValTyrSerGlnLeuThrMetGlnLysAlaPhe 715
QY 1361 CTG 1363
Db 716 Leu 716
RESULT 7
Q9H6K9 PRELIMINARY; PRT; 715 AA.
ID Q9H6K9;
AC Q9H6K9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA: FLJ22171 fis, clone HRC00654 (Epidermal growth factor receptor
DE pathway substrate 8 related protein 2).
GN EPSR2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono T., Sugano S.;
RT "NEO human DNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK025824; BAB15248.1; -
DR EMBL; AY074929; AAL76118.1; -
DR HSSP; Q08509; IAOJ.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00482; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS50002; SH3; 1.
KW Receptor; SH3 domain.
SQ SEQUENCE 715 AA; 80620 MW; DAB0774B04CFEE2 CRC64;
```

```
Alignment Scores:
Pred. No.: 4.39e-38 Length: 715
Score: 584.00 Matches: 163
Percent Similarity: 45.00% Conservative: 71
Best Local Similarity: 31.35% Mismatches: 156
Query Match: 18.68% Indels: 130
DB: 4 Gaps: 19
```

US-09-762-021A-1 (1-1710) x Q9H6K9 (1-715)

```
QY 131 GAGCAGGACGCTATCTATCGAGCCGGGATCCCTCCAGAACAGCCCCACCAGAGGACCCTA 190
Db 190 GluLysIleArgGlnArgGlnSerIleLeuProPro-----11111111
QY 191 GAGCAGAGCTCCCAACCATCCCAAGGCCCCCTGCCA---CGCCACACCACTGCCCCGAGAA 247
```

Db	549	AlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGlyGlnLysTyrTrpGlyPro	568
Qy	1109	-----GAGGCCCTACAGCGGGGACCCCTGGGACCCAGGCCAG-----	1147
Db	569	AlaSerProThrHisLysLeuProProSerPheProGlyAsnLysAspGluLeuMetGln	588
Qy	1148	-----TCACCC	1153
Db	589	HisMetAspGluValAsnAspGluLeuIleArgLysIleSerAsnIleArgAlaGlnPro	608
Qy	1154	TCGCG-----GTTCCAATG-----CTTCGACTTAGCTCG	1183
Db	609	GlnArgHisPheArgValGluArgSerGlnProValSerGlnProLeuThrTyrGluSer	628
Qy	1184	AGGCCTGAAGAGTCACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCACGGTGAGG	1243
Db	629	GlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPheSerProArgIleValGlu	648
Qy	1244	ACACTTGGTGCCTGACGGGGGACCCAGCTACTTCCATTAAGACCTTGGGGAGCTACAGATG	1303
Db	649	AsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLysGluGluLeuLysLys	668
Qy	1304	CTATGCTCACGAGGCGCCACGAATCCCTGTCGCCGGCTGAGGCGTGTCCAGAGGATGCTG	1363
Db	669	ValCysGlyGluGluGlyValArgValTyrSerGlnLeuThrMetGlnLysAlaPheLeu	688
RESULT 8			
Q96BB7	ID	Q96BB7 PRELIMINARY; PRT; 596 AA.	
AC	Q96BB7;		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DE	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	Hypothetical 66.5 kDa protein.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=SKIN;		
RA	Strausberg R.;		
RL	Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: BC015763; AAH15763.1; -		
DR	InterPro: IPR001452; SH3.		
DR	Pfam: PF00018; SH3; 1.		
DR	ProDom: PD000066; SH3; 1.		
DR	ProSITE: PS00002; SH3; 1.		
KW	Hypothetical protein.		
SQ	SEQUENCE 596 AA; 66477 MW; CA417997A8419B0F C		
Alignment Scores:			
Pred. No.:	2.43e-37	Length:	596
Score:	574.50	Matches:	169
Percent Similarity:	44.29%	Conservative:	83
Best Local Similarity:	29.70%	Mismatches:	184
Query Match:	16.37%	Indels:	133
DB:	4	Gaps:	19
US-09-762-021A-1 (1-1710) x Q96BB7 (1-596)			
Qy	2	CGACGCGCTGAAGACCCCTGCAGAGGCTCGGAGGAGAGCTGGACCAAGA---	58
Db	17	AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuHisAsnTyrArgSerGlyArgGly	36
Qy	59	---CCTCGCTTGGAGCGCTTCAGCCAGGCCAGGAC-----AGATGGAGGGGCGCT	106
Db	37	GluArgArgAlaAlaLeuArgAlaThrGlnGluLeuGlnArgAspArgSerPro	56
Qy	107	GCTATGGAAGCGCGCTCCCTATGAGCAGGCGACGCTATCTGGAGCCGGGATCCCTCCA	166
Db	57	AlaAlaGluThrPro---ProLeuGlnArgArg-----ProSer	68

Qy	167	GAAACAGCCCAACGACGAGGCCCTTAGACGACAGCCTC-----CCACCATCCCCAAGG	217
Db	69	VaiArgAlaValIleSerThrValGluArgGlyAlaGlyArgProGlnAlaLys	88
Qy	218	CCCTGCCAGCCCACACCACTGCCCGAGAACCAAGTGCCTTTACTCTGCCTCTCCAAGG	277
Db	89	ProIleProGluAlaGluGluAlaGlnArgProGluProValGlyThrSerSerAsnAla	108
Qy	278	CGGTCTCTTCCCCCGG-----GACCACGACG	304
Db	109	AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu	128
Qy	305	AGGACGAGGAAGTGCTGAACCATGTCCTTAAGGGACATTGAGCTGTTCATGGGAAGCTG	364
Db	129	ArgGluValAspIleLeuAsnHisValPheAspValGluSerPheValserArgLeu	148
Qy	365	GAGAAGGCC-----CAGGCAAGACCACGAGAGAGAANAATTGGGAAAAAAAACAAG	418
Db	149	GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgArgSerArg	168
Qy	419	-----CACGAGGAGTGCTCAC-----CAGSCACG	445
Db	169	ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGluAlaGlu	188
Qy	446	TACATTGACTGCTCCAGAGAATCAAGTACAGCTTCAACCTCTGGGAAGCTGCCAC	505
Db	189	TyrThrAspValLeuGlnLysIleLysTyAlaPheSerLeuLeuAlaArgLeuArgGly	208
Qy	506	TGCTGAGGAGACAGTCCCTGAGCTGTACACATCCCTCTTCAAATCCCTGAATTC	565
Db	209	AsnIleAlaAspProSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet	228
Qy	566	ATCCTGGCCAGGTGCCTGAGGTGCCTAGCAGCCCAAGTCATCTCACCCCTCCTCAC	625
Db	229	IleValasnThrSerglyGlyProGluPheAlaSerSerValArgArgProHisLeuThr	248
Qy	626	CCTAAAGCTATCAACCTGTCTACAGTCTGTCTAAAGCCACCTGAGAGTAACCTTTGGATG	685
Db	249	SerAspAlaValAlaLeuLeuArgaspAsnValThrProArgGluasngluLeuTrpThr	268
Qy	686	GGTTGGGCCACCTGGACCACTAGCCGGGCCGACTGGACAGGGCGATGAGCCCTGCC	745
Db	269	SerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGluGluGlyProPro	288
Qy	746	TACCAACCCACATCTCAGATGACTGGCAACTTCCA-----GAGCCCTCCAGCCACGA	799
Db	289	TyrArgProGluPhePheSerglyrTrpGluProProValThrAspProGlnSerArgAla	308
Qy	800	CCCTTAGGATACCAGGACCTGTTCCTTCGCGGGGAAAGTCATPATGGTAGGAGCAC	859
Db	309	-----TrpGluAspProValGlnLysGlnLeuGlnHisGluArgArgArgGln	325
Qy	860	TCACACTTCTCAGAGAGACACACAACTAT---GACCCTCAGCTGGGAGCCCAAC	916
Db	326	GlnSerAlaProGlnValAlaValAsnGlyHisArgAspLeuGluPro-----	341
Qy	917	TCAGAGCCCTCAGGCCCAAACTCGCCACGACCCCTGAAAATCAAGTCTTG-	970
Db	342	-----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn	359
Qy	971	TACAGTTTTGAAGCTAGGAACCCAGGGAACTGACTGTGTCCAGGAGAGAAGCTGGAG	1030
Db	360	TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu	379
Qy	1031	GTCTTGACACACAGAAGCGGTGGTGGTGTGAAGATGAGCGGGACGAGCGGTAC	1090
Db	380	ValLeuAspAspSerArgLysTrpTrpLysValArgAspProAlaGlyGlnGluGlyTyr	399
Qy	1091	ATTCCAGCAACATCTCGAGGCCCTTACAGCCGGG-----	1126
Db	400	ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisHisSerGlnSer	418
Qy	1127	-----ACCCCTGGG-----	1135

Db	419	ProAlaArgSerLeuAsnSerThrProProProProAlaProAlaProAlaProPro	438
Qy	1135	-----	1135
Db	439	ProAlaLeuAlaArgProArgTTPaspArgProArgTTPaspSerCysAspSerLeuAsn	458
Qy	1135	-----	1135
Db	459	GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluGluLeu	478
Qy	1136	-----ACCAGGGCCAGTCA--CCCTCTCGGGTT-----	1162
Db	479	GlnAlaAArgLeuAlaGlnGlyArgSerGlyProSerArgAlaValProGlyProArgAla	498
Qy	1163	-----CCAATGCTTCGACTTACGTCGAGGCGCTGAAGAGGTCACAGACTGGCTGCAGGCA	1216
Db	499	ProGluProGlnLeuSerProGlySerAspAlaSerGluValArgAlaTrpLeuGlnAla	518
Qy	1217	GAGAACTTCTCCACTGCCACGTGAGGACACTTGGTCCCTGACGGGGAGCCAGCTACTT	1276
Db	519	LysGlyLysPheSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnLeuPhe	538
Qy	1277	CCATAAAGCACTGGGAGCTACGATGCTATGTCACAGAGGAGCCCAAGATCCTGTCC	1336
Db	539	SerLeuGlnLysGluGluLeuArgAlaValSerProGluGluGlyAlaArgValTyrSer	558
Qy	1337	CGCGTGGAGGCTGTCAGAAGGATGCTG	1363
Db	559	GlnValThrValGlnArgSerLeuLeu	567
RESULT 9			
Q8TE68			
ID	Q8TE68	PRELIMINARY;	PRT; 723 AA.
AC	Q8TE68;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Epidermal growth factor receptor pathway substrate 8 related protein		
DE	1		
GN	EPSR1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,		
RA	Romano P., Di Fiore P.P.;		
RT	"Cloning and characterization of novel members of the Eps8 protein		
RT	family";		
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY074928; AAL76117.1; ..		
KW	Receptor.		
SQ	SEQUENCE 723 AA; 80300 MW; 00BCDDFE16F62E13 CRC64;		
Alignment Scores:			
Pred. No.:	3,63e-37	Length:	723
Score:	572.50	Matches:	168
Percent Similarity:	44.22%	Conservative:	84
Best Local Similarity:	29.53%	Mismatches:	184
Query Match:	18.31%	Indels:	133
DB:	4	Gaps:	19
US-09-762-021A-1 (1-1710) x Q8TE68 (1-723)			
Qy	2	GCAGAGCGCACTGCAGACCAACCGCTGCAGAGGGCTCTGGAGGAGAGCTGGAGCAAGA	58
Db	144	AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuHisAsnTyrArgSerGlyArgGly	163
Qy	59	-----CCTCGACTTGGAGGCGCTTCAGCCAGGCCAGGAC-----AGATGGAGGGGGCCT	106
Db	164	GluArgArgAlaAlaLeuArgAlaThrGlnGlnGluLeuGlnArgAspArgSerPro	183


```
QY 107 GCTATGAAGGCCGCTCCCTATGAGCAGCAGCAGCTATCTGGAGCGGGATCCCTCCA 166
Db 184 AlaAlaGluThrPro--ProLeuGlnArgArg-----ProSer 195
QY 167 GAACAGCCCAACAGGACCCCTAGACACAGCTC-----CCACATCCCAAGG 217
Db 196 ValArgAlaValIleSerThrValGluArgGlyAlaGlyArgGlyProGlnAlaLys 215
QY 218 CCCTGCCAGCCACACACAGCTGCCGAGAACCAAGTGCCTTTACTCTGCCTCCTCAAGG 277
Db 216 ProIleProGluAlaGluAlaGlnArgProGluProValGlyThrSerSerAsnAla 235
QY 278 CGCTCTCTTCCCCCGAG-----GACCACAGAG 304
Db 236 AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu 255
QY 305 AGGAGCAGAGTGTGTAACCATGCTTAAGGGACATTGAGCTGTTCATGGGAAGCTG 364
Db 256 ArgGluValAspIleLeuAsnHisValPheAspAspValGluSerPheValSerArgLeu 275
QY 365 GAGAAGGCC-----CAGCAAGACACAGCAGGAAGAAATTTGGGAAAAAACAAG 418
Db 276 GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgSerArg 295
QY 419 -----GACCAGGAGGTCTCAC-----CAGGCACAG 445
Db 296 ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGluAlaGlu 315
QY 446 TACATTGACTGCTTCAGAAAGATCAAGTACAGCTTCAAGCTCTCGGAAGGCTGGCCACC 505
Db 316 TyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAlaArgLeuArgGly 335
QY 506 TGGCTCAAGGAGACAAGTCCCTGAGCTCGTACACATCTCTCAAGTCCCTGAACTTC 565
Db 336 AsnIleAlaAspSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet 355
QY 566 ATCTGCGCCAGGTGCGCTGAGGCTGCGCTAGCAGCCCAAGTATCTCACCCCTCTCCACC 625
Db 356 IleValAsnThrSerGlyGlyProGluPheAlaSerSerValArgArgProHisLeuThr 375
QY 626 CCTAAGCTATCAACTGCTACAGTCTCTTAAGCCCACTGAGAGTAACTTTGGATG 685
Db 376 SerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsnGluLeuTrpThr 395
QY 686 GGGTGGCCGACCGCTGGACCACTAGCCGGCGGACTGGACAGGCGCATGAGCCCTGCC 745
Db 396 SerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGluGlyProPro 415
QY 746 TACCAACCCACATTCAGTACTGGCAACTTCCA-----GAGCCCTCCAGCCCAAGCA 799
Db 416 TyrArgProGluPhePheSerGlyTrpGluProProValThrAspProGlnSerArgAla 435
QY 800 CCCTTAGGATACAGGACCTGTTTCCCTTCGGCGGGGAGTATAGTTAGGAGACACC 859
Db 436 -----TrpGluAspProValGluLysGlnLeuGlnHisGluArgArgArgGln 452
QY 860 TCACACTTTCCTCAGGACAGACACACACCAT--GACCCTCAGCTGGGACCCCAAC 916
Db 453 GlnSerAlaProGluValAlaValAsnGlyHisArgAspLeuGluPro-----468
QY 917 TCCAGGCCCTCAGCCCAACCTGCCAGCCAGCCGCTGAAATGCAAGTCTTG-----970
Db 469 -----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn 486
QY 971 TAGGAGTTTGAAGCTAGGAAACCCAGGAACTGCTGTGTCAGGAGGAGAGAGCTGGAG 1030
Db 487 TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu 506
QY 1031 GTTCTGACACAGCAGCGGTGGTGGTCAACAATCAGCGGCGGACGCGGCTAC 1090
Db 507 ValLeuAspSerArgLysTrpTrpLysValArgAspProAlaGlyGlnGluGlyTyr 526
```

```
QY 1091 ATTCCAAGCAACATCCTGAGCCCTTACAGCCGGG-----1126
Db 527 ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisSerGlnSer 545
QY 1127 -----ACCCTGGG-----1135
Db 546 ProAlaArgSerLeuAsnSerThrProSerProProProAlaProAlaProPro 565
QY 1135 -----1135
Db 566 ProAlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerLeuAsn 585
QY 1135 -----1135
Db 586 GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluGluLeu 605
QY 1136 -----ACCAGGGCCAGTCA---CCCTCTCGGGT-----1162
Db 606 GlnAlaArgLeuAlaGlnGlyArgSerGlyProSerArgAlaValProGlyProArgAla 625
QY 1163 -----CCAATGCTTCGACTTAGCTGAGGCGCTGAAGAGGTACAGACTGGCTGCAGGCA 1216
Db 626 ProGluProGlnLeuSerProGlySerAspAlaSerGluValArgAlaTrpLeuGlnAla 645
QY 1217 GAGAACTTCTCCACTCCACGCTGAGGACACTTGGCTCCCTGACGGGAGCCAGCTACTT 1276
Db 646 LysGlyPheSerSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnLeuPhe 665
QY 1277 CCATTAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCCAGCAATCTCTCC 1336
Db 666 SerLeuGlnArgGluGluLeuArgAlaValSerProGluGluGlyAlaArgValTyrSer 685
QY 1337 CGCTGAGGCTGTCAAGAGGATGCTG 1363
Db 686 GlnValThrValGlnArgSerLeuLeu 694
RESULT 10
Q9NXH0 PRELIMINARY; PRT; 596 AA.
AC Q9NXH0:
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE CDNA FLJ20258 fis, clone COLF7250.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Oiyashiki M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigaki T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK00265; BAA91041.1; -.
DR HSP; Q08509; IAOJ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 596 AA; 66517 MW; 3A586087464F7812 CRC64;
Alignment Scores:
Pred. No.: 4, 21e-37 Length: 596
Score: 571.50 Matches: 169
Percent Similarity: 44.23% Conservative: 83
Best Local Similarity: 29.70% Mismatches: 184
Query Match: 18.28% Indels: 133
```


DB: 4 Gaps: 19
US-09-762-021a-1 (1-1710) x QNKH0 (1-596)
QY 2 GCAGAGCCACTCAACAGCAGCTCGCAGAGGCTCTGGAGGAGAGCTCGGACCAAGA---- 58
||||| : : : : : |||||
Db 17 AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuHisAsnTyrArgSerGlyArgGly 36
QY 59 ---CCTGACTTGGAGGCTTCAGCCAGCCAGGAC-----AGATGGAGGGGCT 106
||||| : : : : : |||||
Db 37 GluArgArgAlaAlaAlaLeuArgAlaThrGlnGluLeuGlnArgAspArgSerPro 56
QY 107 GCTATGGAAGCCGCTCCCTATGAGCAGCAGCAGCTATCTGGAGCCGGGATCCCTCCA 166
||||| ||| : : : : : : : : : :
Db 57 AlaAlaGluThrPro---ProLeuGlnArg-----ProSer 68
QY 167 GNACAGCCCCCAGCAGCAGCCTAGACACAGCCTC-----CCACCATCCCAAG 217
||||| : : : : : |||||
Db 69 ValArgAlaValIleSerThrValGluArgGlyAlaGlyArgGlyArgProGlnAlaLys 88
QY 218 CCCTGCGCCAGCCACACAGCTGCCGAGACCAAGTGCCTTACTCTGCCTCCTCCAAG 277
||||| : : : : : |||||
Db 89 ProIleProGluAlaGluGlnAlaGlnArgProGluProValGlyThrSerSerAsnAla 108
QY 278 CGTCTCTCTCCCGGAG-----GACCCAGAG 304
||||| : : : : : |||||
Db 109 AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu 128
QY 305 AGGAGCAGGAGTCTCAACCATGCTCTAAGGACATTTAGCTGTTCTATGGGAAGCTG 364
||||| : : : : : |||||
Db 129 ArgGluValAspIleLeuAsnHisValPheAspValGluSerPheValSerArgLeu 148
QY 365 GAGAAGGCC-----CAGGCAAAAGACCAGCAGGAAGAATAATTGGGAAAAAACAAG 418
||||| : : : : : |||||
Db 149 GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgArgSerArg 168
QY 419 -----GACCAGGAGGTCTCACC-----CAGCAGCAG 445
||||| : : : : : |||||
Db 169 ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGluAlaGlu 188
QY 446 TACATTGACTCTCCCAAGATCAAGTACAGCTTCAACCTCCCTGGGAAGCTGGCCACC 505
||||| : : : : : |||||
Db 189 TyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAlaArgLeuArgGly 208
QY 506 TGGTGAAGAGACAAAGTCCCTGAGCTCGTACACATCTCTCAAGTCCCTGAACCTC 565
||||| : : : : : |||||
Db 209 AsnIleAlaAspProSerSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet 228
QY 566 ATCTGGCCAGTGCCCTGAGCTGGCTAGCAGCCCAAGTGATCTACCCCTCCTCACC 625
||||| : : : : : |||||
Db 229 IleValAsnThrSerGlyProGluPheAlaSerSerValArgArgProHisLeuThr 248
QY 626 CCTAAAGCTATCAACCTCTACAGTCCTGTCTAAGCCCACTGAGAGTAACCTTTGGATG 685
||||| : : : : : |||||
Db 249 SerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsnGluLeuThrPthr 268
QY 686 GGGTTGGCCAGCTGGACCACTAGCCGGCCGCTGAGCAGCGCATGAGCCCTGCC 745
||||| : : : : : |||||
Db 269 SerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGluGluGlyProPro 288
QY 746 TACCAACCCACATCTCAGATGACTGGCAACTTCCA-----GAGCCCTCCAGCCCAAGCA 799
||||| : : : : : |||||
Db 289 TyrArgProGluPhePheSerGlyTrpGluProProValThrAspProGlnSerArgAla 308
QY 800 CCTTAGGATACCAGGACCTGTTTCCCTTCGGCGGGGAAGTCATAGTTAGGAGCACC 859
||||| : : : : : |||||
Db 309 -----TrpGluAspProValGluLysGlnLeuGlnHisGluArgArgArgGln 325
QY 860 TCACACTTCTCTCAGGAGAAGACACAAACCAT---GACCCTCAGCCTGGGGACCCCAAC 916
||||| : : : : : |||||
Db 326 GlnSerAlaProGlnValAlaValaLysnGlyHisArgAspLeuGluPro----- 341
QY 917 TCCAGGCCCTCCAGCCCAACCTGCCAGCAGCCCTGAAATGCAAGTCTTG----- 970

Db 342 -----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn 359
||||| : : : : : |||||
QY 971 TACAGATTGAAGTAGAACCCAGGAACTGACTGTGTCTCCAGGAGAGAACTGGAG 1030
||||| : : : : : |||||
Db 360 TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu 379
QY 1031 GTTCTGGACCAACAGCGGTGGTGGTGAAGTAATGAGGGGACGAGCGGCTAC 1090
||||| : : : : : |||||
Db 380 ValLeuAspSerArgLysTrpTrpLysValArgAspProAlaGlyGlnGluGlyTyr 399
QY 1091 ATCCAAAGCAACATCTGGAGCCCTACAGCCGGG----- 1126
||||| : : : : : |||||
Db 400 ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisHisSerGlnSer 418
QY 1127 -----ACCCCTGG----- 1135
||||| : : : : : |||||
Db 419 ProAlaArgSerLeuAsnSerThrProProProProAlaProAlaProAlaProPro 438
QY 1135 ----- 1135
Db 439 ProAlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerLeuAsn 458
QY 1135 ----- 1135
Db 459 GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluGluLeu 478
QY 1136 -----ACCCAGGCCAGTCA---CCCTCTCGGTT----- 1162
||||| : : : : : |||||
Db 479 GlnAlaArgLeuAlaGlnGlyArgSerGlyProSerArgAlaValProGlyProArgAla 498
QY 1163 -----CCAATGCTTCGACTTAGCTCGAGGCTGAGAGGCTCACAGCTGGCTGCAGCA 1216
||||| : : : : : |||||
Db 499 ProGluProGlnLeuSerProGlySerAspAlaSerGluValArgAlaTrpLeuGlnAla 518
QY 1217 GAGAACTTCTCCACTGCCGCTGAGGACACTTGGTCCCTGACGGGAGCCAGCTACTT 1276
: : : : : ||||| : : : : : |||||
Db 519 LysGlyPheSerSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnLeuPhe 538
QY 1277 CGCATAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCCAATCCTGTCC 1336
: : : : : ||||| : : : : : |||||
Db 539 SerLeuGlnArgGluGluLeuArgAlaValSerProGluGluGlyAlaArgValTrpSer 558
QY 1337 CGGTGGAGGCTGTCAGAGGATGCTG 1363
: : : : : ||| : : : |||
Db 559 GlnValIleValGlnArgSerLeuLeu 567
RESULT 11
Q9D2M6 PRELIMINARY; PRT; 652 AA.
AC Q9D2M6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 4632407K17RIK protein.
GN 4632407K17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=SKIN;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL: AK019490; BAB31756.1; -.
 DR HSSP: Q08509; IAOJ.
 DR MGD: MGI:1914675; 4632407K17Rik.
 DR InterPro: IPR001452; SH3.
 DR Pfam: PF00018; SH3; 1.
 DR ProDom: PD000066; SH3; 1.
 DR SMART: SM00326; SH3; 1.
 DR PROSITE: PS50002; SH3; 1.
 KW SH3 domain.
 SQ SEQUENCE 652 AA; 73007 MW; D43DAA59C2D7C638 CRC64;

Alignment Scores:

Pred. No.: 2,04e-36 Length: 652
 Score: 563.00 Matches: 162
 Percent Similarity: 42.81% Conservative: 79
 Best Local Similarity: 28.77% Mismatches: 194
 Query Match: 18.00% Indels: 128
 DB: 11 Gaps: 15

US-09-762-021A-1 (1-1710) x Q9D2M6 (1-652)

QY 2 GCAGCGGACTGAAGACCGCTGCAGAGCGTCTCGGAGGAGAGCTGGAGCAAGA---- 58
 DB 80 AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuGlnAsnTyrArgSerGlyArgGly 99
 QY 59 ---CCTCGACTTGGAGCCCTTACAGCGCCAGGACAGATGGAGG-----GGGCGT 106
 DB 100 GluArgArgAlaAlaAlaLeuArgAlaThrGlnGluLeuArgGlyAlaSerPro 119
 QY 107 GCTATGGAAGCGCTGCTATGAGACGACGACGCTATCTGGAGCGGGATCCCTCCA 166
 DB 120 AlaAlaGluThrPro--ProLeuGlnArg----- 128
 QY 167 GAACGCCCCACAGGACCTTAGACGACGCTCCACCATCCCA---AGGCCCTGTG 223
 DB 129 ---ArgProSerValArgLeuValIleAsnThrValGluProSerAlaValArgGlyArg 147
 QY 224 CCAGCCACACAGTGCCTGAGCAACCAAGTGCCTTTACTCTGCTCTCCAGGCGGTCC 283
 DB 148 ProGlnValGluSerIleProGluThrGluGluAlaArgLysProAspGlnAlaArgThr 167
 QY 284 TCTTCCCCGAGGACCA----- 301
 DB 168 ThrSerSerAlaAspProThrSerProAspLeuGlyProArgGlyProGluLeuAlaGly 187
 QY 302 -----GAGAGGACGAGGAGTCTGAACCATGCTCTAAGGACATGAGCTGTTC 352
 DB 188 LeuGlnAlaGluArgAspValAspIleLeuAsnHisValPheAspValGluSerPhe 207
 QY 353 ATGGGAAGCTGAGGAAGGCCCGCAAGAC-----AGCAGGAG 394
 DB 208 ValSerArgLeuGlnLysSerAlaGluAlaThrArgValLeuGluHisArgGluArgGly 227
 QY 395 AAGAATTTGGGAAAAAACAAGACGACGAGGAGGTCTC----- 433
 DB 228 ArgArgThrArgArgAlaAlaGlyGluGlyLeuThrThrLeuArgAlaLysProPro 247
 QY 434 ACCGAGCAGATACACTGCTCTCCAGACATCAAGTACAGCTTCAACCTCTCGGGA 493
 DB 248 ThrGluAlaGluThrThrAspValLeuGlnLysIleLysIleAsnThrValPheSerLeuAla 267

QY 494 AGGCTGGCCACCTGGCTGAAGGAGACAAAGTCCCTGAGCTCGTACACATCTCTTTCAG 553
 DB 268 ArgLeuArgGlyAsnIleAlaAsnProSerSerProGluLeuHisPheLeuGly 287
 QY 554 TCCCTGAACCTTCTATCTGCGCCAGGTCCCTGAGGTGGCTAGCAGCCCAAGTGATCTCA 613
 DB 288 ProLeuGlnMetIleValAsnThrSerGlyGlyProGluPheAlaLysSerValArgArg 307
 QY 614 CCCCTCTCACCCCTAAAGCTATCAACCTGTACAGCTCTGTCTAAGCCCACTGAGAT 673
 DB 308 ProHisLeuThrLeuGluAlaValThrLeuLeuArgAspAsnValThrProGlyGluAsn 327
 QY 674 AACCTTTGGATGGGTGGGCCAGCCCTGACCACTAGCCGGGCCAGCTGGACAGCGAT 733
 DB 328 GluLeuThrThrSerLeuGlyAspSerTrpThrCysProGlyValGluLeuProGlu 347
 QY 734 GAGCCCTGCTCTACCAACCCACATTCAGATGACTGCGCACTCCAGAGCCCTCCAGC 793
 DB 348 GluGlySerProTyrSerProGluPheTyrAsnGlyTrp-----GluProProAla 364
 QY 794 CAAGCACCTTAGGA-----TACCAGGACCTGTGTTCCCTTCGGGGGGAAGTATAGG 847
 DB 365 ThrAspProGlnGlyArgProTrpGluAspProValGluGlnLeuHisGluLys 384
 QY 848 TTAGGAGCACCTCACACTTCTCAGGAGAAAGACACACACCACTGACCTCAGCCTGG 907
 DB 385 ArgArgArgGlnGlnSerAlaProGlnValAlaValAsnGlyGlnGln----- 400
 QY 908 GACCCCACTCCAGCCCTCAGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTGC 967
 DB 401 AspProGluLeuGluThrGluSerGlnLeuGluGlyLysAlaArgLysTrpValLeuCys 420
 QY 968 TTGTAGGAGCTTAGAGCTAGGAACCCAGCGGAACTGACTGTGTCCAGGAGAGAGCTG 1027
 DB 421 AsnTyrAspPheGlnAlaArgAsnGlySerGluLeuSerValLysHisArgAspValLeu 440
 QY 1028 GAGGTTCGAGCACAGCAAGCGGTGGTGGTGAAGAATGAGCGGAGGAGCGGC 1087
 DB 441 GluValLeuAspArgArgLysTrpTrpLysValArgAspHisGlnGlyGlnGly 460
 QY 1088 TACATTCGAAGCAACATCTCTGAGGCC----- 1114
 DB 461 TyrValProTyrAsnIleLeuThrProHisProGlyProGlnValHisArgSerGlnSer 480
 QY 1115 -----CTACAGCGCGGACCCCTGG----- 1135
 DB 481 ProAlaArgHisLeuGluThrSerThrProProProProAlaProAlaPro 500
 QY 1135 ----- 1135
 DB 501 ThrGlnValArgProGlnTrpAspSerCysAspSerLeuAsnSerLeuAspProSerGlu 520
 QY 1136 -----ACCCAG 1141
 DB 521 LysGluLysPheSerGlnMetLeuCysValAsnGluLeuGlnSerArgLeuAlaGln 540
 QY 1142 GGCCAGCTCA---CCCTCTCGGTT-----CCAATGCTTCGA 1174
 DB 541 GlyArgSerGlyProSerArgValThrProGlyProArgAlaGlnGluProGlnLeuSer 560
 QY 1175 CTTAGCTCGAGCCCTGAAGAGTCTACAGACTGGCTGCGAGCAGAGAACTTCCACTGCC 1234
 DB 561 ProArgSerGluAlaSerValValArgAlaTrpLeuGlnThrLysGlyPheSerSerGly 580
 QY 1235 AGGCTGAGCACACTTGGTCCCTGAGCGGAGCCCACTACTTCGCATACACCTGGGAG 1294
 DB 581 ThrValGluAlaLeuGlyValLeuThrGlyAlaGlnLeuPheSerLeuGlnLysGluGlu 600
 QY 1295 CTACAGATGCTGTCCAGGAGCGGCCCACTCTCTCGCTGGAGGTGTGTCAG 1354
 DB 601 LeuArgAlaValCysProGluGluGlyAlaArgValTyrSerGlnValThrValGlnArg 620
 QY 1355 AGGATGCTG 1363

[illegible]


```

Db 577 IleGluArgGlyGlyGlyAlaAlaGlyHisGlyHisGlyHisGlyProGly 596
QY 872 -----CAGGAGAACACACACAAACATGACCT 898
Db 597 AlaGlyProThrLeuSerAlaIleThrAlaGlyLeuGlnAsnLeuHisThrArgGluSer 616
QY 899 CAGCTGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCT----- 940
Db 617 ArgSerGlyAsnGlyTyrGlyAlaGlyAlaGlyProGlyProSerSerGluLeuGlyGly 636
QY 940 ----- 940
Db 637 GlyArgGlyLeuProAsnValSerAspAspGlnMetLeuGluSerTrpLeuGluAsp 656
QY 941 -----GCCAGCCAGCCGTAAGTCTGTGACGAGTTTGAAGCTAGGAACCCA 994
Db 657 LeuGlnAlaThrGlyAlaLysIleValLeuValThrTyrProArgThrAlaAsnAsp 676
QY 995 CGGGAAGTCTGTGTCAGGAGAGAGAGTGGAGTTCTGGACACAGCAAGCGGTGG 1054
Db 677 LysGluLeuSerValMetArgGlyGluTyrLeuGluIleLeuAspThrArgLysTrp 696
QY 1055 TGGCTGGTGAAGATGAGCGGGAGGAGCGGCTACATTCACGAACATCTGGAGCCC 1114
Db 697 TrpLysAlaArgAsnMetArgGlyGlnValAlaHisValProHisThrIleValThrPro 716
QY 1115 CTA-----CAGCCGGGGACC 1129
Db 717 PheAsnPheGlyAspGlyAlaGlnPheTyrGlyGlnGlnGlnProThr 736
QY 1130 -----CCTGGGAGCCAGCGGCCAG 1147
Db 737 GlyProThrGlyProGlyAsnLysSerArgSerGlyAspAsnProGlyMetGluGlnArg 756
QY 1148 TCACCTCTCGGGTTCCTCAATGCTTCCA 1174
Db 757 SerProAspProThrAspMetMetArg 765

RESULT 14
Q90329 ID Q90329 PRELIMINARY: PRT; 778 AA.
AC Q90329;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE AROUSER.
GN ARU OR CG4276.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu C.H., Bourne J., Harris S.J., O'Kane C.J., Moffat K.G.;
RT "Drosophila epidermal growth factor receptor kinase substrate Eps8.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF208262; AAF21013.1; -.
DR HSSP; Q08509; IAOJ.
DR FlyBase; FBgn0029095; aru.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain. 778 AA; 85132 MW; E0ED1485F5285189 CRC64;
SQ SEQUENCE

```

Alignment Scores:

```

Pred. No.: 1.42e-23 Length: 778
Score: 402.00 Matches: 522
Percent Similarity: 37.31% Conservative: 153
Best Local Similarity: 26.01% Mismatches: 122
Query Match: 12.86% Indels: 172
DB: 5 Gaps: 13

US-09-762-021A-1 (1-1710) x Q90329 (1-778)
QY 281 TCCTCTTCCCGAGGACCCAGAGGAGGAGTGCTGAACCATGCTCCTCAAGGAC 340
Db 298 SerThrSerSerGluLysTyrGluArgAspValAlaValLeuAsnHisCysPheAsp 317
QY 341 APTGAGCTGTTTCAAGGAAAGCTGGAGAGGAGCCAGGCA-----AAG 382
Db 318 IleGluLysPheIleAlaArgLeuGlnHisAlaAlaAlaSerArgGluLeuGluArg 337
QY 383 ACCACGAGAGAGAAATTTGGGAAAAAACAAGGACCCAGGAGGTCTCACC----- 436
Db 338 ArgArgArgAsnArgLysSerLysLysArgAspProGlyGluGlyLeuLeuThrLeuArg 357
QY 437 -----CAGGCACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTC 481
Db 358 ThrArgProProHisGluLysGluPheValAspIlePheAlaLysPheLysLeuSerPhe 377
QY 482 AACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTCCCTCAGCTCGTACAC 541
Db 378 AsnLeuLeuAlaLysLeuLysAlaHisIleHisAspProAsnAlaProGluLeuValHis 397
QY 542 ATCTCTTCAAGTCCCTGAACCTTCATCTCCGCGCCAGGTGCGCCT-----GAGGCT 589
Db 398 PheLeuPheThrProLeuAlaLeuIleValGluAlaSerSerAspThrTyrTyrGluSer 417
QY 590 GGCCTAGGACGCCAAGTATCTCACCCTCTCACCCTCAACCTAAAGCTATCAACCTGCTACAG 649
Db 418 GlnLeuProAlaArgValValAsnProLeuLeuThrArgGluAlaIleAsnLeuLeuIle 437
QY 650 TCCTGCTAAAGCCACCTGAGAGTAACCTTTGGTGGGTGGGCCCGCCCTGGACCACT 709
Db 438 AsnCysValThrSerLysGluThrGluLeuThrArgSerLeuGlyAspAlaTrpValIle 457
QY 710 AGCCGGCCGACTGGACGCGCATGAGCCCTGCCCTACCAACCCACATTCACAGATGAC 769
Db 458 ProArgAspGlnTrpLys---AspAspValGlySerTyrHisProValLeuAspGly 476
QY 770 TGG-----CAACTTCAGAGCCCTCCAGCCCAACCCCTTA 805
Db 477 TrpSerProAspTyrIleAsnAspGluLeuGluProProAsnSerProAla 496
QY 805 ----- 805
Db 497 HisValSerLysArgArgLeuGluValGlnAlaGlyProGlyLeuAsnGlyArgGlyGly 516
QY 806 ---GGATACAGGACCCCTGTTTC----- 826
Db 517 GlyGlyTyrAspAspTyrAspSerGlyAsnGlyMetAsnMetAlaMetGlyIleGluLys 536
QY 827 -----CTTCGG-----CGGGAGAGTCA 844
Db 537 TyrThrIleHisHisGlyAsnAspGlyLeuArgGluArgGluArgGluArgArgAsp 556
QY 845 AGGTTTAGGAGCACCTCA----- 862
Db 557 ArgAlaGlyAlaIleSerAlaSerAspPheAsnAlaArgSerGluLeuSerPheAspSer 576
QY 863 -----CATTCTCTCT--- 871
Db 577 IleGluGlyArgGlyGlyAlaAlaGlyHisGlyHisGlyHisGlyHisGlyProGly 596
QY 872 -----CAGGAGAACACACACATGACCT 898
Db 597 AlaGlyProThrLeuSerAlaIleThrAlaGlyLeuGlnAsnLeuHisThrArgGluSer 616

```

```

QY 899 CAGCCTGGGAGCCCAACATCCAGGCCCTCCAGGCCCAACCT----- 940
Db 617 ArgSerGlyAsnGlyTyrGlyAlaGlyAlaGlyProGlyProSerSerGluLeuGlyGly 636
QY 940 ----- 940
Db 637 GlyGlyArgGlyLeuProAsnValSerAspAspGlnMetLeuGluSerTrpLeuGluAsp 656
QY 941 -----GCCAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCA 994
Db 657 LeuGlnAlaThrGlyAlaLysIleValLeuValThrTyrProArgThrAlaAsnAsp 676
QY 995 CGGAAGTACTGTGTCCAGGAGAGAAAGTGGAGTTCTGGACACAGCAAGCGGTGG 1054
Db 677 LysGluLeuSerValMetArgGlyGlyLeuGluIleLeuAspAspThrArgLysTrp 696
QY 1055 TGCCTGGTGAAGATGAGCGGACGCGGCTACATCCCAAGCAACATCTCGGAGCCC 1114
Db 697 TrpLysAlaArgAsnMetArgGlyGlnValAlaHisValProHisThrIleValThrPro 716
QY 1115 CTA-----CAGCGCGGGGACC 1129
Db 717 PheAsnPheGlyAspGlyAlaGlnPheTyrGlyGlnGlnGlnProProThr 736
QY 1130 -----CCTGGGACCCAGGCCAG 1147
Db 737 GlyProThrGlyProGlyAsnLysSerArgSerGlyAspAsnProGlyMetGluGlnArg 756
QY 1148 TCACCCCTCTCGGTTCCATGCTTCA 1174
Db 757 SerProAspProThrAspMetMetArg 765

RESULT 15
Q9GZ02
ID Q9GZ02 PRELIMINARY; PRT; 437 AA.
AC Q9GZ02;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE DRC3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99445952; PubMed=10514543;
RA Wu K., Xu Z., Wang M., Xu X., Han Y., Cao Y., Wang R., Sun Y., Wu M.;
RT "Cloning and expression analyses of down-regulated cDNA C6-2A in human
RL esophageal cancer.";
RL Chung-Hua I Hsueh I Chuan Hsueh Tsa Chih 16:325-327(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Wang M., Chen B., Wu K., Xu X., Han Y., Cai Y., Wang J., Xu Z., Wu M.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF282168; AAG03039.1; -.
DR EMBL; AF282167; AAG03038.1; -.
DR HSP; Q08509; IAOJ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF000018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 437 AA; 48898 MW; C52F003A6B3491EE CRC64;

Alignment Scores:
Pred. No.: 2,22e-23 Length: 437
Score: 399.00 Matches: 119
Percent Similarity: 46.228 Conservative: 52
Best Local Similarity: 32.16% Mismatches: 132
Query Match: 12.76% Indels: 67

```

```

US-09-762-021a-1 (1-1710) x Q9GZ02 (1-437)
QY 434 ACCAGGCACAGTACATTCAGCTTCCAGAACATCAAGTACAGCTTCAACCTCCTGGGA 493
Db 59 SerGluAlaGluTyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAla 78
QY 494 AGCGTGGCCACCTGGCTGAAGGAGACAGTGGCCCTGAGCTCGTACACATCTCTTCTCAAG 553
Db 79 ArgLeuArgGlyAsnIleAlaAspProSerSerProGluLeuLeuHisPheLeuPheGly 98
QY 554 TCCTGAAGTTCATCCTCGCCAGGTGCTGAGGTGGCTAGCAGCCCAAGTGTCTCA 613
Db 99 ProLeuGlnMetIleValAsnThrSerGlyGlyProGluPheAlaSerValArgArg 118
QY 614 CCCCTCTCACCCTAAAGCTATCAACCTGCTACAGTCTCTAGCCACCTGAGAGT 673
Db 119 ProHisLeuThrSerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsn 138
QY 674 AACCTTTGGATGGGTGGCCAGCCTGACCTAGCCGCCGCTGACGACGCGCAT 733
Db 139 GluLeuTrpThrSerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGlu 158
QY 734 GAGCCCTCCCTACCAACCCACATTTCTCAGATGACTGGCAACTTCCA-----GAGCCC 787
Db 159 GluGlyProProTyrArgProGluPhePheSerGlyTrpGluProProValThrAspPro 178
QY 788 TCCAGCCCAAGCACCCTTAGGATACAGGACCCCTGTTTCCCTTCGG----- 832
Db 179 GlnSerArgAla-----TrpGluAspProValGluGlnLeuGlnHisGluArg 195
QY 833 -----CGG 835
Db 196 ArgArgArgGlnValThrGlnAlaThrGlnGlnArgGlyTrpGluValArgGlyArg 215
QY 836 GGAAGTCAT-----AGGTAGGAGACACTCACACTTCTCCTCAGGAGAGACACAC 886
Db 216 GlyArgSerAlaTrpProArgLeuThrArgLeuSerTyrPheLeuGlnGlnSerAla--- 234
QY 887 AACCATGACCTCAGCCTGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCACG 946
Db 235 -----ProGlnValAlaValAsnGlyHisArgAspLeuGluProGlu---SerGlu 250
QY 947 CCAGCCCTGAAATGCAAGTCTTG-----TACGAGTTTGAAGCT 985
Db 251 ProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsnTyrAspPheGlnAla 270
QY 986 AGAACCCAGGGAAGTACTGTGTCCAGGAGAGAAAGTGGAGTCTTGGACACACAGC 1045
Db 271 ArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGluValLeuAspAspSer 290
QY 1046 AAGCGTGTGTGCTGAAGATGAGCGGACGAGCGGCTACATTCACCAACATC 1105
Db 291 ArgLysTrpTrpLysValArgAspProAlaGlyGlnGluGlyTrpValProTyrAsnIle 310
QY 1106 CTGGAGCCCTTACAGCGGGGACCCCTGGGACCCAGCCAGTACACCTCTCGGTTCCA 1165
Db 311 LeuThrPro---TyrProGlyProArgLeuHisHisSerGlnSerProAlaArg---- 327
QY 1166 ATGCTTCGACTTAGCTCAGGCGCTGAAGAGGTGACAGACTGGCTGAGGAGAGAACTTC 1225
Db 328 -----SerLeuAsnSerThrProProProAlaProAlaProAlaProProPro 345
QY 1226 TCCACTGCCACGGTGAGG-----ACACTTGGTCCCTGACGGGG 1264
Db 346 AlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerGluAsnGly 365
QY 1265 AGCCAGCTACTTCGCATTAAGACCTGGGAGCTACAGATGCTATGTCACAGAGGAGGCCCA 1324
Db 366 -----LeuAspProSerGlu-----LysGluLysPhePro 375
QY 1325 CGAATCCTCTCCCGGCTGGAGGCTGTCAAGAGATGCTGGGGATAGCCCTTAGGCACCA 1384

```

```

Db  376 ArgCysSerSerThrArgAsnCysArgArg-AlaTyrProArgAlaAlaArgAspAr 395
QY  1385 GCTTAGACACCTCCAGAACCCAGGCCCC 1412
Db  395 gAlaAlaGlnSerGlnGlyProAlaPro 404
```

Search completed: February 25, 2003, 19:43:01
Job time : 73 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 07:32:52 ; Search time 1688 Seconds

(without alignments)

16406.565 Million cell updates/sec

Title: US-09-762-021A-1

Perfect score: 1710

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	662	38.7	662	14	BM824107
2	600.8	35.1	604	14	BM744431
3	587.4	34.4	589	14	BM742903
4	560.4	32.8	562	14	BM756251
5	547.4	32.0	549	14	BM770581
6	542.4	31.7	556	10	AW778776

7	530.4	31.0	532	14	BM769901
8	517.6	30.3	504	14	BM741405
c	496.4	29.0	507	12	BF001871
	491.8	28.8	551	10	AW841643
11	477.2	27.9	483	14	BM797626
12	473.2	27.7	478	14	BM792701
c	472	27.6	534	10	AW937714
c	454.4	26.6	456	9	AI339104
15	443.6	25.9	901	14	BQ951422
16	417	24.4	521	14	BM824199
17	413.8	24.2	431	14	BQ320003
c	409.4	23.9	411	9	AA861042
19	404.4	23.6	417	14	BM757009
c	395	23.1	427	14	BM818771
c	389	22.7	401	9	AA134985
21	378.6	22.1	728	13	BI284392
c	374.8	21.9	378	9	AA868144
c	366.6	21.4	773	13	BG966755
c	356.6	20.9	685	9	AI746552
26	354.6	20.7	419	14	BQ326654
c	352.2	20.6	422	13	BG958438
28	344.6	20.2	512	9	AA710419
29	341.6	20.0	449	13	BM445779
c	332.2	19.4	356	12	BF755318
c	331.6	19.4	672	10	AV378666
c	325.8	19.1	413	9	AA134946
c	322.4	18.9	329	13	BG982594
c	321.2	18.8	606	17	AO022003
c	319.2	18.7	369	9	AI626100
c	312.4	18.3	385	14	BQ320515
37	309.4	18.1	322	13	BG982587
c	307.6	18.0	468	12	BF334476
c	305.2	17.8	310	9	AA922724
c	302.2	17.3	524	9	AA535447
41	279.8	16.4	509	9	AI429197
42	274.2	16.0	362	9	AA135023
43	265.6	15.5	368	10	AW951951
44	263.4	15.4	271	10	AW818756
45	263.4	15.4	271	10	AW818837

ALIGNMENTS

RESULT 1	BM824107	BM824107	662 bp	mrna	linear	EST 06-MAR-2002
LOCUS	K-EST0095517	S22SNU16n1	Homo sapiens	cdna	clone S22SNU16n1-89-A05	
DEFINITION	5', mRNA sequence.					
ACCESSION	BM824107					
VERSION	BM824107.1	GI:19180520				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.N., Park,H.S., Kim,S. and Kim,Y.S.					
TITLE	1 (bases 1 to 662)					
JOURNAL	21C Frontier Korean EST Project 2001					
COMMENT	Unpublished (2002)					
	Contact: Kim YS					
	Genome Research Center					
	Korea Research Institute of Bioscience & Biotechnology					
	52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea					
	Tel: +82-42-860-4470					
	Fax: +82-42-860-4409					
	Email: yongsung@mail.kribb.re.kr					
	Plate: 89 row: A column: 05					
	High quality sequence stop: 662.					
FEATURES	Location/Qualifiers					
source	1. .662					

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNU16n1-89-A05"
/clone_lib="S22SNU16n1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/note="Organ: Stomach; Vector: pT7n3-Pac; Site_1: EcoRI;
Site_2: NotI; The S22SNU16 library was contributed by the
Soares laboratory and it was constructed as described by the
Bonafido, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

BASE COUNT 152 a 210 c 190 g 110 t
ORIGIN
Query Match 38.7%; Score 662; DB 14; Length 662;
Best Local Similarity 100.0%; Pred. No. 2.1e-155; Indels 0; Gaps 0;
Matches 662; Conservative 0; Mismatches 0;
QY 679 TTGGATGGGTTGGGCCAGCCTGGACCACACTAGCCGGGCGGACTGGACAGGGGATGAGCC 738
Db 1 TTGGATGGGTTGGGCCAGCCTGGACCACACTAGCCGGGCGGACTGGACAGGGGATGAGCC 60
QY 739 CTGGCCCTACCAACCCACACATTTCTAGATGACTGGCAACTTCCAGAGCCCTCCAGCCAAAGC 798
Db 61 CTGGCCCTACCAACCCACACATTTCTAGATGACTGGCAACTTCCAGAGCCCTCCAGCCAAAGC 120
QY 799 ACCCTTAGGATACAGGACCCCTTTCCCTTCGGCGGGGAAGTCATAGTTAGGGAGCAGC 858
Db 121 ACCCTTAGGATACAGGACCCCTTTCCCTTCGGCGGGGAAGTCATAGTTAGGGAGCAGC 180
QY 859 CTCACACTTTCTTCAGGGAAGACACACACCACTAGCCCTCAGCCCTGGGAGCCCCCAACTC 918
Db 181 CTCACACTTTCTTCAGGGAAGACACACACCACTAGCCCTCAGCCCTGGGAGCCCCCAACTC 240
QY 919 CAGGCCCTCAGCCCCCAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTGTACGAGTT 978
Db 241 CAGGCCCTCAGCCCCCAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTGTACGAGTT 300
QY 979 TGAAGCTAGGAACCCACGGGAAGTGTGTGTCAGGGAGAGAGCTGGAGTTCTGGA 1038
Db 301 TGAAGCTAGGAACCCACGGGAAGTGTGTGTCAGGGAGAGAGCTGGAGTTCTGGA 360
QY 1039 CCACAGCAAGCGGTGGTGGCTGGTGAAGATAGGCGGAGCGGCGGTACATTTCCAAG 1098
Db 361 CCACAGCAAGCGGTGGTGGCTGGTGAAGATAGGCGGAGCGGCGGTACATTTCCAAG 420
QY 1099 CAACATCTGGAGCCCTACAGCGGGGGAGCCCTGGGAGCCAGGGCCAGTCCCTCTCG 1158
Db 421 CAACATCTGGAGCCCTACAGCGGGGGAGCCCTGGGAGCCAGGGCCAGTCCCTCTCG 480
QY 1159 GGTTCGAATGCTTCGACTAGCTCGAGGCTGAAGAGGTCACAGCTGGGTCGAGCAGA 1218
Db 481 GGTTCGAATGCTTCGACTAGCTCGAGGCTGAAGAGGTCACAGCTGGGTCGAGCAGA 540
QY 1219 GAATTTCTCCATGCGCACCGGTGAGGACACTTGGGTCCCTCAGCGGGAGCGAGTACTTCG 1278
Db 541 GAATTTCTCCATGCGCACCGGTGAGGACACTTGGGTCCCTCAGCGGGAGCGAGTACTTCG 600
QY 1279 CATAAGACCTGGGGAGCTACAGATGCTATGTCCAGGAGGGCCCCACGAATCTGTCCCG 1338
Db 601 CATAAGACCTGGGGAGCTACAGATGCTATGTCCAGGAGGGCCCCACGAATCTGTCCCG 660
QY 1339 GC 1340
Db 661 GC 662

RESULT 2
BM744431
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM744431 604 bp mRNA linear EST 01-MAR-2002
K-EST001978 S6SNU620 Homo sapiens cDNA clone S6SNU620-15-H03 5',
mRNA sequence.
BM744431
BM744431.1 GI:19065760
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 604)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-gong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 15 row: H column: 03
High quality sequence stop: 604.

FEATURES
source

1..604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S6SNU620-15-H03"
/clone_lib="S6SNU620"
/sex="F"
/tissue_type="Ascites"
/cell_type="Scattering floating"
/cell_line="SNU-620"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then deapped
with tabacco acid pyrophosphatase (TAP). The deapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transformation of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 147 a 205 c 142 g 110 t
ORIGIN
Query Match 35.1%; Score 600.8; DB 14; Length 604;
Best Local Similarity 99.7%; Pred. No. 4.9e-140; Indels 0; Gaps 0;
Matches 602; Conservative 0; Mismatches 2;
QY 411 AAAACAAGGACACAGGAGGCTCCACCCAGGCACAGTACATTGACTCTCCAGAAGATCA 470
Db 1 AAAACAAGGACACAGGAGGCTCCACCCAGGCACAGTACATTGACTCTCCAGAAGATCA 60
QY 471 AGTACAGCTTCAACCTCTCTGGGAAGGCTGCCACCTGGCTGGAGGAGACAAGTGCCTCG 530
Db 61 AGTACAGCTTCAACCTCTCTGGGAAGGCTGCCACCTGGCTGGAGGAGACAAGTGCCTCG 120
QY 531 AGCTCGTACACATCTCTTCAAGTCCCTCAAGTTCATCTCGCCAGGTGCCCTGAGGCTG 590
Db 121 AGCTCGTACACATCTCTTCAAGTCCCTCAAGTTCATCTCGCCAGGTGCCCTGAGGCTG 180

```
QY 591 GCCTAGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 650
|||||
Db 181 GCCTAGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 240
QY 651 CCTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 710
|||||
Db 241 CCTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 300
QY 711 GCCGGGCCACTGGACGGCGATGAGCCCTCGCCCTACCAACCCACATCTCAGATGACT 770
|||||
Db 301 GCCGGGCCACTGGACGGCGATGAGCCCTCGCCCTACCAACCCACATCTCAGATGACT 360
QY 771 GGCAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAGGACCTGTTCCTTC 830
|||||
Db 361 GGCAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAGGACCTGTTCCTTC 420
QY 831 GCCGGGGAGTCATAGTTAGGAGCACCTCACACTTTCCTCAGGAGAGACACACAACC 890
Db 421 GCCGGGGAGTCATAGTTAGGAGCACCTCACACTTTCCTCAGGAGAGACACACAACC 480
QY 891 ATGACCCCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTGCCAGCCAG 950
|||||
Db 481 ATGACCCCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTGCCAGCCAG 540
QY 951 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAACCTGCTGTGG 1010
|||||
Db 541 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAACCTGCTGTGG 600
QY 1011 TCCA 1014
Db 601 TCCA 604

RESULT 3
LOCUS BM742903
DEFINITION K-EST0015941 S6SNUG20 Homo sapiens cDNA clone S6SNUG20-10-H03 5',
mRNA sequence.
ACCESSION BM742903
VERSION BM742903.1 GI:19064232
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 589)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 10 row: H column: 03
High quality sequence stop: 589.
Location/Qualifiers
1. 589
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S6SNUG20-10-H03"
/clone_lib="S6SNUG20"
/sex="F"
/tissue_type="Ascites"
/cell_type="Scattering floating"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
```

```
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 144 a 201 c 139 g 105 t
ORIGIN

Query Match 34.4%; Score 587.4; DB 14; Length 589;
Best Local Similarity 99.8%; Pred. No. 1.le-136;
Matches 588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 411 AAACAAGGACCCAGGGAGGCTCTACCCAGGACAGTACATTGACTGCTTCCAGAGATCA 470
|||||
Db 1 AAACAAGGACCCAGGGAGGCTCTACCCAGGACAGTACATTGACTGCTTCCAGAGATCA 60
QY 471 AGTACAGCTTCAACCTCCTGGGAAGGCTGGCCACTGGCTGAAGGAGACAAGTGCCCTG 530
|||||
Db 61 AGTACAGCTTCAACCTCCTGGGAAGGCTGGCCACTGGCTGAAGGAGACAAGTGCCCTG 120
QY 531 AGCTCGTACACATCCTCTTCAAGTCCCTGAACTTATCCTTGCCAGGTGCCCTGAGGCTG 590
|||||
Db 121 AGCTCGTACACATCCTCTTCAAGTCCCTGAACTTATCCTTGCCAGGTGCCCTGAGGCTG 180
QY 591 GCCTAGAGCCCAAGTGTATCTACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGT 650
|||||
Db 181 GCCTAGAGCCCAAGTGTATCTACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGT 240
QY 651 CTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 710
Db 241 CTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 300
QY 711 GCCGGGCCACTGGACGGCGATGAGCCCTCGCCCTACCAACCCACATCTCAGATGACT 770
Db 301 GCCGGGCCACTGGACGGCGATGAGCCCTCGCCCTACCAACCCACATCTCAGATGACT 360
QY 771 GGCAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAGGACCTGTTCCTTC 830
Db 361 GGCAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAGGACCTGTTCCTTC 420
QY 831 GCCGGGGAGTCATAGTTAGGAGCACCTCACACTTTCCTCAGGAGAGACACACAACC 890
Db 421 GCCGGGGAGTCATAGTTAGGAGCACCTCACACTTTCCTCAGGAGAGACACACAACC 480
QY 891 ATGACCCCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTGCCAGCCAG 950
Db 481 ATGACCCCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTGCCAGCCAG 540
QY 951 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAACCTGCTGTGG 999
Db 541 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAACCTGCTGTGG 589

RESULT 4
LOCUS BM756251
DEFINITION K-EST0034532 S6SNUG20 Homo sapiens cDNA clone S6SNUG20-28-D01 5',
mRNA sequence.
ACCESSION BM756251
VERSION BM756251.1 GI:19085866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 562)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.krabb.re.kr
Plate: 28 row: D column: 01
High quality sequence stop: 562.
Location/Qualifiers
1..562

FEATURES
source

1..562
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S6SNU620-28-D01"
/clone_lib="S6SNU620"
/sex="F"
/tissue_type="Ascites"
/cell_type="Scattering floating"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."
136 a 195 c 130 g 101 t

BASE COUNT
ORIGIN

Query Match 32.8%; Score 560.4; DB 14; Length 562;
Best Local Similarity 99.8%; Pred. No. 6.8e-130;
Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 411 AAAACAAGCAGCAGGAGGCTCACCCAGGCACAGTACATTGCTTCCAGAGATCA 470
|||||
Db 1 AAACAAGCAGCAGGAGGCTCACCCAGGCACAGTACATTGCTTCCAGAGATCA 60
|||||
QY 471 AGTACAGCTTCAACCTCTCGGAAGGCTGCCACCTGGCTGAAGGAGACAACTGGCCCTG 530
|||||
Db 61 AGTACAGCTTCAACCTCTCGGAAGGCTGCCACCTGGCTGAAGGAGACAACTGGCCCTG 120
|||||
QY 531 AGTCTGTACATCTCTTCAAGTCTCCTGAAGTCTATCTGCGCAGGTGCCCTGAGGCTG 590
|||||
Db 121 AGCTGTGTACATCTCTTCAAGTCTCCTGAAGTCTATCTGCGCAGGTGCCCTGAGGCTG 180
|||||
QY 591 GCTGTACGCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGT 650
|||||
Db 181 GCTGTACGCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGT 240
|||||
QY 651 CTGTCTTAAGCCCACTGTAGAGTAACCTTTGGATGGGGTTGGCCCGCAGCTTGACCACTA 710
|||||
Db 241 CTGTCTTAAGCCCACTGTAGAGTAACCTTTGGATGGGGTTGGCCCGCAGCTTGACCACTA 300
|||||
QY 711 GCGGGGCCACTGGACGGGATGAGCCCTTGCCTTACCACCCACATTTCTCAGATGACT 770
|||||

Db 301 GCCGGCCGACTGGACAGCGGATGAGCCCTGCCCTTACCACCCACATTTCTCGGATGACT 360
QY 771 GCGAACTTCCAGAGCCCTCCAGCCCAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCTTC 830
|||||
Db 361 GCGAACTTCCAGAGCCCTCCAGCCCAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCTTC 420
|||||
QY 831 GCGGGGGAAGTCATAGTTTAGGGAGCACCTCACACTTTCTCAGGAGAGACACACAACC 890
|||||
Db 421 GCGGGGGAAGTCATAGTTTAGGGAGCACCTCACACTTTCTCAGGAGAGACACACAACC 480
|||||
QY 891 ATGACCCCTCAGCTGGGGACCCCAACTCCAGGCCCTCCAGCCCCCAACCTGCCAGCCAG 950
|||||
Db 481 ATGACCCCTCAGCTGGGGACCCCAACTCCAGGCCCTCCAGCCCCCAACCTGCCAGCCAG 540
|||||
QY 951 CCCTGAAATGCAAGTCTTGTA 972
|||||
Db 541 CCCTGAAATGCAAGTCTTGTA 562
|||||

RESULT 5
BM770581
LOCUS
DEFINITION
5', mRNA sequence.
549 bp mRNA linear EST 04-MAR-2002
K-EST0054176 S6SNU620s1 Homo sapiens cDNA clone S6SNU620s1-16-H08
BM770581
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BM770581
BM770581.1 GI:19100196
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 549)
Oh,K.J., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.
21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.krabb.re.kr
Plate: 16 row: H column: 08
High quality sequence stop: 549.
Location/Qualifiers
1..549
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S6SNU620s1-16-H08"
/clone_lib="S6SNU620s1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Scattering floating"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library. After analyzing and

sequencing about 2,000 - 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(dT)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized cDNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the substracted cDNA libraries were constructed by transfection of the remaining DNA into competent cells E. coli Top10F' with electroporation method.

BASE COUNT 133 a 193 c 127 g 96 t
ORIGIN

Query Match 32.0%; Score 547.4; DB 14; Length 549;
Best Local Similarity 99.8%; Pred. No. 1.3e-126;
Matches 548; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 411 AAAACAAGACAGGAGGCTCTACCCAGGCACAGTACATTCAGTCTCCAGAGATCA 470
|||||
Db 1 AAAACAAGACAGGAGGCTCTACCCAGGCACAGTACATTCAGTCTCCAGAGATCA 60
Qy 471 AGTACAGCTTCAACCTCTCGGAAGCTGGCCACCTGCTGAAGGAGACAAGTGCCCTG 530
|||||
Db 61 AGTACAGCTTCAACCTCTCGGAAGCTGGCCACCTGCTGAAGGAGACAAGTGCCCTG 120
Qy 531 AGCTGTACATCTCTTCAAGTCCCTGAAGTCCCTGACCTTCACTTCTGCGCCAGGTGCCCTGAGGCTG 590
|||||
Db 121 AGCTGTACATCTCTTCAAGTCCCTGAAGTCCCTGACCTTCACTTCTGCGCCAGGTGCCCTGAGGCTG 180
Qy 591 GCCTAGCAGCCCAAGTGTATCTCACCCCTCTCACCCCTTAAAGCTATCAACCTGTCTACAGT 650
Db 181 GCCTAGCAGCCCAAGTGTATCTCACCCCTCTCACCCCTTAAAGCTATCAACCTGTCTACAGT 240
Qy 651 CTTGTCTAAGCCCACTGACAGTAACTTTGGATGGGTTGGGCCAGCCTGGACCACTA 710
Db 241 CTTGTCTAAGCCCACTGACAGTAACTTTGGATGGGTTGGGCCAGCCTGGACCACTA 300
Qy 711 GCCGGCCGACTGGACAGGCATGAGCCCTGCTTACCAACCCACATCTTCAGATGACT 770
Db 301 GCCGGCCGACTGGACAGGCATGAGCCCTGCTTACCAACCCACATCTTCAGATGACT 360
Qy 771 GCGAACTTCCAGAGCCCTCCAGCAAGCACCTTAGGATACAGGACCTGTGTTCCCTTC 830
Db 361 GCGAACTTCCAGAGCCCTCCAGCAAGCACCTTAGGATACAGGACCTGTGTTCCCTTC 420
Qy 831 GCGGGGAAGTCATAGTTAGGAGCACCTTACACATTTTCTTCAGGAGAAGACACACAACC 890
Db 421 GCGGGGAAGTCATAGTTAGGAGCACCTTACACATTTTCTTCAGGAGAAGACACACAACC 480
Qy 891 ATGACCTTCAGCTGGGAGCCCAACTCCAGGCCCTCAGGCCCCAAACCTGCCAGCCAG 950
Db 481 ATGACCTTCAGCTGGGAGCCCAACTCCAGGCCCTCAGGCCCCAAACCTGCCAGCCAG 540
Qy 951 CCCTGAAAA 959
Db 541 CCCTGAAAA 549

RESULT 6
AW78776/c
LOCUS AW78776 556 bp mRNA linear EST 12-MAY-2000
DEFINITION hol2h03.x1 NCI_CGAP_col4 Homo sapiens cDNA clone IMAGE:3037205 3' similar to SW:EP58_MOUSE Q08509 EPIDERMAL GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8. ; mRNA sequence.
ACCESSION AW78776
VERSION AW78776.1 GI:7793366
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 413.
Location/Qualifiers 1..556
FEATURES
source /organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3037205"
/clone_lib="NCI_CGAP.Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-Sport6; Site:1: Salt: Site_2: NotI: Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT 117 a 131 c 168 g 140 t
ORIGIN

Query Match 31.7%; Score 542.4; DB 10; Length 556;
Best Local Similarity 99.8%; Pred. No. 2.3e-125;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1151 CCCTCTCGGGTTCGAATGCTTCGACTTAGCTCGAGCCCTGAAGAGGTACAGACTGGCTG 1210
Db 544 CCCTCTCGGGTTCGAATGCTTCGACTTAGCTCGAGCCCTGAAGAGGTACAGACTGGCTG 485
Qy 1211 CAGGAGAGAATCTTCCACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAG 1270
Db 484 CAGGAGAGAATCTTCCACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAG 425
Qy 1271 CTACTTCCATTAAGACCTTGGGGAGCTACAGATGCTATGTCCACAGAGGCCCCAGGAATC 1330
Db 424 CTACTTCCATTAAGACCTTGGGGAGCTACAGATGCTATGTCCACAGAGGCCCCAGGAATC 365
Qy 1331 CTGTCCCGGCTGGAGGCTGTCAGAAGGATGCTGGGGATAAGCCCTTAGGCACCACTTAG 1390
Db 364 CTGTCCCGGCTGGAGGCTGTCAGAAGGATGCTGGGGATAAGCCCTTAGGCACCACTTAG 305
Qy 1391 ACACCTCCAGAAGACAGGCCCGCTGTATGCAAGATGGCAGATGATATACCCATTAGAGCC 1450
Db 304 ACACCTCCAGAAGACAGGCCCGCTGTATGCAAGATGGCAGATGATATACCCATTAGAGCC 245
Qy 1451 CCGAGAATTCCTTCTTGATCCAGTTTGGCAGCAAAACCCACACCCAGCTCAGACAGC 1510
Db 244 CCGAGAATTCCTTCTTGATCCAGTTTGGCAGCAAAACCCACACCCAGCTCAGACAGC 185
Qy 1511 AAAACAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGCTGTGTGGAG 1570
Db 184 AAAACAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGCTGTGTGGAG 125
Qy 1571 CCTCCCCAGTAAACCACTTATTTATTTACCTCTTTTCCCAACCTGGAGCATTTATGCGCTA 1630
Db 124 CCTCCCCAGTAAACCACTTATTTATTTACCTCTTTTCCCAACCTGGAGCATTTATGCGCTA 65
Qy 1631 GGCTTGTCAGAAATCTGTTTCAGTCCCTCTCTTCTCAATAAAAGCATCTTCAAGCTTGTGA 1690
Db 64 GGCTTGTCAGAAATCTGTTTCAGTCCCTCTCTTCTCAATAAAAGCATCTTCAAGCTTGTGA 5

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

QY 1691 AAAA 1694
|||||
Db 4 AAAA 1

RESULT 7

BM769901

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

Source

BM769901 532 bp mRNA linear EST 04-MAR-2002
K-EST0053246 S14K402 Homo sapiens cDNA clone S14K402-24-B06 5',
mRNA sequence.

BM769901

BM769901.1 GI:19099516

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 532)

Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,

Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and

Kim, Y.S.

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 24 row: B column: 06

High quality sequence stop: 532.

Location/Qualifiers

1..532

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S14K402-24-B06"

/clone_lib="S14K402"

/cell_line="K402"

/lab_host="Top10F"

/note="Organ: Stomach; Vector: pTZ18Rpl; Site_1: EcoRI;

Site_2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of T4 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was

adjusted to have about 60nt. The cDNA vector was

circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transfection of

competent cells E. coli Top10F by electroporation method.

The cDNA libraries constructed by this method are

full-length enriched cDNA library."

BASE COUNT 123 a 187 c 125 g 97 t

ORIGIN

Query Match 31.0%; Score 530.4; DB 14; Length 532;

Best Local Similarity 99.8%; Pred. No. 2.3e-122;

Matches 531; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 475 CACGTTCAACCTCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTGCCTGAGCT 534

|||||

Db 1 CACGTTCAACCTCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTGCCTGAGCT 60

|||||

QY 535 CGTACATCTCTTCAAGTCCCTGAACCTTCATCTTGGCCAGGTGCGCTGGCTGGCT 594

|||||

Db 61 CGTACATCTCTTCAAGTCCCTGAACCTTCATCTTGGCCAGGTGCGCTGGCTGGCT 120

|||||

QY 595 AGAGCCCAAGTATCTACCCCTCTCTACCCCTTAAGCTATCAACCTGCTACAGTCCTG 654

|||||

Db 121 AGCAGCCCAAGTGATCTCACCCCTCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCCTG 180
QY 655 TCTAAGCCACCTGAGAGTAACCTTTTGGATGGGTTGGGCCCCAGCTGGACCACTAGCG 714
|||||
Db 181 TCTAAGCCACCTGAGAGTAACCTTTTGGATGGGTTGGGCCCCAGCTGGACCACTAGCG 240
QY 715 GCGCAGCTGGACAGCGATGAGCCCTGCCCTTACCAACCCACATTTCTCAGATGACTGGCA 774
Db 241 GCGCAGCTGGACAGCGATGAGCCCTGCCCTTACCAACCCACATTTCTCAGATGACTGGCA 300
QY 775 ACTTCCAGAGCCCTCCAGCAAGCACCCTTAGGATACCAAGACCCCTGTTTCCCTTCGGCG 834
Db 301 ACTTCCAGAGCCCTCCAGCAAGCACCCTTAGGATACCAAGACCCCTGTTTCCCTTCGGCG 360
QY 835 GGAAGTCTATAGTTAGGAGCACCCTCACACCTTCTCAGGAGAGACACACAACCATCA 894
Db 361 GGAAGTCTATAGTTAGGAGCACCCTCACACCTTCTCAGGAGAGACACACAACCATCA 420
QY 895 CCTCAGCCTGGGACCCCAACTCCAGGCGCTCCAGCCCAACCAACCTGCCAGCCAGCCCT 954
Db 421 CCTCAGCCTGGGACCCCAACTCCAGGCGCTCCAGCCCAACCAACCTGCCAGCCAGCCCT 480
QY 955 GAAATGCAAGTCTTGTACGAGCTTTGAAGCTAGGAACCCACGGGAACCTGACT 1006
|||||
Db 481 GAAATGCAAGTCTTGTACGAGCTTTGAAGCTAGGAACCCACGGGAACCTGACT 532

RESULT 8
BM741405 604 bp mRNA linear EST 01-MAR-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

K-EST0014022 S6SNU620 Homo sapiens cDNA clone S6SNU620-7-B10 5',
mRNA sequence.

BM741405

BM741405.1 GI:19062734

EST.

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

1 (bases 1 to 604)

Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,

Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and

Kim, Y.S.

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 7 row: B column: 10

High quality sequence stop: 604.

Location/Qualifiers

1..604

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S6SNU620-7-B10"

/clone_lib="S6SNU620"

/sex="F"

/tissue_type="Ascites"

/cell_type="Scattering floating"

/cell_line="SNU-620"

/lab_host="Top10F"

/note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;

Site_2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of T4 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was

adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transfection of competent cells E. coli Top10F⁺ by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT	147 a	207 c	142 g	108 t
ORIGIN				
Query Match	30.38;	Score 517.6;	DB 14;	Length 604;
Best Local Similarity	97.58;	Pred. No. 3.9e-119;		
Matches 589;	Conservative 0;	Mismatches 9;	Indels 6;	Gaps 6;
Qy	410	AAAAAAGAGGAGGCTTCAACCCAGGACACAGTACATTGACTGCTTCCAGAAGATC	469	
Db	1	AAAAAAGGAGGAGGAGGCTTCAACCCAGGACACAGTACATTGACTGCTTCCAGAAGATC	60	
Qy	470	AAGTACAGCTTCAACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCCT	529	
Db	61	GAGTACAGCTTCAACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCCT	120	
Qy	530	GAG-CTCGTACACATCCT-CTTCAAGTCCCT-GAACTTCATCCT-GGCAGAGTGCCCTG-	584	
Db	121	GAGACTCGTACACATCCTCGCTTCAAGTCCCTAGAATTCATCCTGGGCGAGGTGCCCTGC	180	
Qy	585	AGGCTGGCTAGCAGCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGC	644	
Db	181	AGGCTGGCTAGCAGCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGC	240	
Qy	645	TACAGTCTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGA	704	
Db	241	TACAGTCTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGCA	300	
Qy	705	CAACTAG-CCGGCGGACTGGACAGGCGATGAGCCCTTGGCCCTACCAACCCACATFTCA	763	
Db	301	CAACTAGTCCGGCGGACTGGACAGGCGATGAGCCCTTGGCCCTACCAACCCACATFTCG	360	
Qy	764	GATGACTGGCAACTTCCAGAGCCCTCCAGCAAGCACCTTAGGATACCAAGACCCCTGT	823	
Db	361	GATGACTGGCAACTTCCAGAGCCCTCCAGCAAGCACCTTAGGATACCAAGACCCCTGT	420	
Qy	824	TCCCTTCGGCGGGAGTCATAGGTTAGGAGCACTTCCCTCAGGAGAGACA	883	
Db	421	TCCCTTCGGCGGGAGTCATAGGTTAGGAGCACTTCCCTCAGGAGAGACA	480	
Qy	884	CACACCATGACCTCAGCCTGGGAGCCCAACTCCAGGCCCTCCAGCCCAACCTGCC	943	
Db	481	CACACCATGACCTCAGCCTGGGAGCCCAACTCCAGGCCCTCCAGCCCAACCTGCC	540	
Qy	944	CAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGAAGT	1003	
Db	541	CAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGAAGT	600	
Qy	1004	ACTG 1007		
Db	601	ACTG 604		

RESULT 9
BF001871/c
LOCUS
DEFINITION
7994f11.x1 NCI-CCAP Col6 Homo sapiens cDNA clone IMAGE:3314157 3,
similar to SW:EP8_MOUSE Q08509 EPIDERMAL GROWTH FACTOR RECEPTOR
KINASE SUBSTRATE EPS8. ; mRNA sequence.
ACCESSION
BF001871
VERSION
BF001871.1
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 507)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 468.
Location/Qualifiers
1. 507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3314157"
/clone_lib="NCI-CCAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI-CCAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 108 a 121 c 149 g 128 t 1 others
ORIGIN

Query Match	29.08;	Score 496.4;	DB 12;	Length 507;
Best Local Similarity	98.68;	Pred. No. 8e-114;		
Matches 500;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
Qy	1181	TCGAGGCTGAAGAGTTCACAGACTGGCTGCAGGAGAGAACTTCTCCACTGCCACGGTG	1240	
Db	507	TCGAGGCTGAAGAGTTCACAGACTGGTGCAGGAGAGAACTTCTCCACTGCCACGGTG	448	
Qy	1241	AGGACACTTGGTCTCCACGGGAGCCAGCTACTTCCATAAGACCTGGGAGGTACAG	1300	
Db	447	AGGACACTTGGTCTCCACGGGAGCCAGCTACTTCCATAAGACCTGGGAGGTACAG	388	
Qy	1301	ATGCTATGTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1360	
Db	387	ATGCTATGTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	328	
Qy	1361	CTGGGATTAAGCCCTTAGGCACACAGCTTAGACACCTCAAGAACAGGAGGAGGAGG	1420	
Db	327	CTGGGATTAAGCCCTTAGGCACACAGCTTAGACACCTCAAGAACAGGAGGAGGAGG	268	
Qy	1421	AAGATGCGAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTTCGATCCCAAGTTG	1480	
Db	267	AAGATGCGAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTTCGATCCCAAGTTG	208	
Qy	1481	CAGCAAAACCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1540	
Db	207	CAGCAAAACCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	148	
Qy	1541	CAACAGTGTCCCTTCTGGCTGTGTGGAGCTCCCAAGTAACCACTTATTTATTTACC	1600	
Db	147	CAACAGTGTCCCTTCTGGCTGTGTGGAGCTCCCAAGTAACCACTTATTTATTTACC	88	
Qy	1601	TCCTTCCCAACCTGGAGCATTTATGCTAGGCTTGTCAAGAACTCTGTTCAGTCCCTC	1660	
Db	87	TCCTTCCCAACCTGGAGCATTTATGCTAGGCTTGTCAAGAACTCTGTTCAGTCCCTC	28	

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
|||||
Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

RESULT 10

AW841643
LOCUS RC1-CN0017-120200-012-a09 CN0017 Homo sapiens linear EST 18-MAY-2000
DEFINITION
ACCESSION AW841643
VERSION AW841643.1 GI:7935626
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 551)
Dias Neto,E., Garcia Correa,R., Verjovskij-Almelda,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.P.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

MEDLINE

20202663

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=6t2-RC1-CN0017-120200-012-a09&t3=2000-02-12&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 25
High quality sequence stop: 551.

FEATURES

source

1..551
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CN0017"
/dev_stage="Adult"

/note="Organ: colon_normal; Vector: puc18; Site_1: Smal;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORFESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
145 a 163 c 164 g 79 t

BASE COUNT

ORIGIN

Query Match 28.8%; Score 491.8; DB 10; Length 551;
Best Local Similarity 98.9%; Pred. No. 1.2e-112;
Matches 517; Conservative 0; Mismatches 2; Indels 4; Gaps 2;

QY 19 CAGCCTGCAGAGGCTCTGGAGGAGAGAGCTGGAGCAA---AGACCTCGACTTGGAGGCT 75

|||||
Db 29 CAGCCTGCAGAGGCTCTGGAGGAGAGAGCTGGAGCAAAGCAGCCTCGACTTGGAGGCT 88

|||||
QY 76 TCAGCCAGGCCAGGACAGATGAGGGGGCC-TGCTATGAAAGGCGCTCCCTATGGAGC 134

|||||
Db 89 TCAGCCAGGCCAGGACAGATGAGGGGGCCATGCTGTGGAAAGGCGCTCCCTATGGAGC 148

|||||
QY 135 AGGCAGCTATCTGGAGCGGGGGATCCCTCCAGACAGAGCCCCCAGCAGGAGCCCTAGAGC 194

|||||
Db 149 AGGCAGCTATCTGGAGCGGGGGATCCCTCCAGACAGAGCCCCCAGCAGGAGCCCTAGAGC 208

QY 195 ACAGCTCCCAACCATCCCAAGGCCCTGCCACGCCACACCAAGTCCCGAGAACCAAGTG 254
|||||
Db 209 ACAGCTCCCAACCATCCCAAGGCCCTGCCACGCCACACCAAGTCCCGAGAACCAAGTG 268
|||||
QY 255 CTTTACTCTCGCTCTCAAGGGGGTCTCTTCCCGAGGAGCCAGAGGAGGAGG 314
|||||
Db 269 CTTTACTCTCGCTCTCAAGGGGGTCTCTTCCCGAGGAGCCAGAGGAGGAGGAGG 328
|||||
QY 315 AAGTCTCAACCATCTCTTAAGGGACATTGAGCTGTTTCATGGAAAGCTGGAGAAGGCC 374
|||||
Db 329 AAGTCTCAACCATCTCTTAAGGGACATTGAGCTGTTTCATGGAAAGCTGGAGAAGGCC 388
|||||
QY 375 AGGCAAGACAGCAGGAGGAAGAAATTTGGGAAAAAAGACAGGAGGAGGAGTCTCA 434
|||||
Db 389 AGGCAAGACAGCAGGAGGAAGAAATTTGGGAAAAAAGACAGGAGGAGGAGTCTCA 448
|||||
QY 435 CCCAGGCACAGTACATTGACTGCTTCCAGAGAGATCAAGTACAGCTTCAACCTCTCGGAA 494
|||||
Db 449 CCCAGGCACAGTACATTGACTGCTTCCAGAGAGATCAAGTACAGCTTCAACCTCTCGGAA 508
|||||
QY 495 GGCTGGCCACCTGGCTGAAGGAGAGCAAGTGCCCTGAGCTCGT 537
|||||
Db 509 GGCTGGCCACCTGGCTGAAGGAGAGCAAGTGCCCTGAGCTCGT 551
|||||

RESULT 11

BM797626
LOCUS K-EST0080876 S22SNU16n1 Homo sapiens cDNA clone S22SNU16n1-76-C12
DEFINITION 5', mRNA sequence.
ACCESSION BM797626
VERSION BM797626.1 GI:19145858
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 483)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.

REFERENCE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoseun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 76 row: C column: 12
High quality sequence stop: 483.

Location/Qualifiers
1..483
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S22SNU16n1-76-C12"
/clone_lib="S22SNU16n1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"

/note="Organ: Stomach; Vector: pT7T3-Pac; Site_1: EcoRI;
Site_2: NotI; The S22SNU16 library was contributed by the
Soares laboratory and it was constructed as described by
Bonaldio, M.F., Lennon, G. and Soares, M.B. (1996), Genome
Research 6(9): 791-806. RNA was prepared from harvested
cells of SNU-16 culture. SNU-16 cell was obtained from
Korean Cell Line Bank (KCLB). SNU-16 was established from
ascitic fluids of Korean patients by Park J.G. et al.
(1990), Cancer Res 50: 2773-2780."

FEATURES

source

FEATURES
source

JOURNAL	Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
MEDLINE	M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
COMMENT	Simpson,A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed
	sequence tags
	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
	20202663
	Contact: Simpson A.J.G.
	Laboratory of Cancer Genetics
	Ludwig Institute for Cancer Research
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
	Brazil
	Tel: +55-11-2704922
	Fax: +55-11-2707001
	Email: asimpson@ludwig.org.br
	This sequence was derived from the FAPESP/LICR Human Cancer Genome
	Project. This entry can be seen in the following URL
	(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-QV3-DT0045-221)
	299-046-a016t3-1999-12-22&t4=1)
	Seq primer: puc 18 forward
	High quality sequence start: 16
	High quality sequence stop: 365.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

```

1. 331
/organism="Homo sapiens"
/db_xref="taxon:9608"
/clone_lib="DT0045"
/dev_stage="Adult"
/note="Organ: denis drash; Vector: puc18; Site_1: SmaI;
Site_2: SnaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT      96 a   143 c   162 g   133 t
ORIGIN

Query Match      27.6%   Score 472;   DB 10;   Length 534;
Best Local Similarity 97.2%   Pred. No. 1.le-107;
Matches 513; Conservative 0; Mismatches 10; Indels 5; Gaps 3;

```

Qy	249	CAAGTGCCTTTACTCTCGCTCCTCCAAAGGGGGTCTCTTTCCCCCGAGAGCCAGCAGAGGG	308
Db	534	CAAGTGCCTGTACTCTCGCTCCTCGAAGGGGGTCTCTTTCCCGCGAGGATCCAGAGAGGG	475
Qy	309	ACGAGGAAGTCTGAACCATCTCTTAAGGGACATTGAGCTGCTTATGGGAAAGCTGGAGA	368
Db	474	ACGAGGAAGTCTGAACCATCTCTTAAGGGACATTGAGCTGCTTATGGGNAAGCTGGAGA	415
Qy	369	AGGCCCAAGCAAGACCCAGCAGGAAGAAGAAATTTGGGAAAAA - - - ACAAAGGACCAGG	425
Db	414	AGGCCCAAGCAAGACCCAGCAGGAAGAAGAAATTTGGGAAAGAAGACAAGAACCCAGG	355
Qy	426	GAGGCTCACCCAGGCACAGTACA - TTGACTGCTTCAGAGATCAAGTACAGCTTCAAC	484
Db	354	GAGGCTCACCCAGGCACAGTACAGTTGACTGCTTCAGAAGAATCAAGCACAGCTTCAAC	295
Qy	485	CTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCTTGAGCTGTTACACATC	544
Db	294	CTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCTTGAGCTGTTACACATC	235
Qy	545	CTCTTCAAGTCCCTGAACTTTCATCCTGGCCAGGTGCCCTGAGGCTGGCCTAGCAGCCAA	604
Db	234	CTCTTCAAGTCCCTGAACTTTCATCCTGGCCAGGTGCCCTGAGGCTGGCCTAGCAGCCAA	175
Qy	605	GTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACTGCTTACAGTCTCTGTCTAAGCCCA	664
Db	174	GTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACTGCTTACAGTCTCTGTCTAAGCTCA	115
Qy	665	CCTGAGAGTAACCTTTGGATGGGGTTGGGGCCAGCCCTGGACCACTAGCCGGGCGACTGG	724
Db	114	CCTGAGAGTAACCTTTGGATGGGGTTGGGGCCAGCCCTGGACCACTAGCCGGGCGACTGG	55

Qy	725	ACAGCGATGAGCCCTGCCCTACCACCCACATTCTCAGTACTGG	772
Dd	54	ACAGCGATGAG-CCCTGCCCTACCACCCACATTCTCAGACGCAGGS	8
RESULT 14			
AJ39104/c			
LOCUS	AJ39104	456 bp mRNA linear	EST 13-FEB-1999
DEFINITION	G106all.x1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1946780 3'		
ACCESSION	AJ39104	mRNA sequence.	
VERSION	AJ39104.1	GI:4076031	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 456) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
TITLE	Contact: Robert Strausberg, Ph.D. Email: cgapbs@email.nih.gov Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M. Bento Soares, Ph.D. DNA sequencing by: Greg Lennon, Ph.D. Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/bbrp/image/image.html		
JOURNAL	Insert length: 606 Std Error: 0.00		
COMMENT	Seq primer: -40UP from Gibco High quality sequence stop: 447.		

[illegible]

```

1. .456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1946780"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from 3 pooled
germ cell tumors, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
98 a 105 c 137 g 116 t

```

[illegible]

	Query Match	26.6%	Score 454.4;	DB 9;	Length 456;
	Best Local Similarity	99.8%;	Pred. No. 2.8e-103;		
	Matches 455;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1236	CGGTGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATTAAGACCTGGGGAGC	1295		
Db	456	CGGTGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATTAAGACCTGGGGAGC	397		
Qy	1296	TACAGATGCTATGTCCACAGGAGGCCCCACGAATCTGTCCCGGCTGGAGGCTGTCAAG	1355		
Db	396	TACAGATGCTATGTCCACAGGAGGCCCCACGAATCTGTCCCGGCTGGAGGCTGTCAAG	337		
Qy	1356	GGATGCTGGGGATTAAGCCCTTTAGGCACAGCTTTAGACACCTTCCAAGAACACAGGCCCGCT	1415		
Db	336	GGATGCTGGGGATTAAGCCCTTTAGGCACAGCTTTAGACACCTTCCAAGAACACAGGCCCGCT	277		
Qy	1416	GATGCAAGATGGCAGATCTGTATACCCCATTTAGAGCGCCGAGAAATTCCTCTTCTGGATCCCA	1475		

Db 276 GATGAAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATCTCTTCTGATGCCA 217

Qy 1476 GTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAAATGAGAGGCCAGAGGC 1535

Db 216 GTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAAATGAGAGGCCAGAGGC 157

Qy 1536 TGAAGCAAAACAGTGTCCCTCTGCTGTGTGGCTGTGTGGAGCCCTCCCGAGTAACCACTATTATT 1595

Db 156 TGAAGCAAAACAGTGTCCCTCTGCTGTGTGGAGCCCTCCCGAGTAACCACTATTATT 97

Qy 1596 TTACCTCTTCCCAACCTGGAGCAATTTATGCTAGGCTTGTCAAGATCTGTTCACTCC 1655

Db 96 TTACCTCTTCCCAACCTGGAGCAATTTATGCTAGGCTTGTCAAGATCTGTTCACTCC 37

Qy 1656 CTCTCTCTCTCAATAAAGCATCTTCAAGCTTGTAA 1691

Db 36 CTCTCTCTCTCAATAAAGCATCTTCAAGCTTGA 1

RESULT 15

BO951422

LOCUS

DEFINITION AGENCOURT_8836792 NCI_CGAP_Co24 Mus musculus cDNA clone

IMAGE:6397884 5', mRNA sequence.

BO951422

BO951422.1 GI:22365900

EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cdapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: L1AM13896 row: h column: 13
High quality sequence stop: 651.
Location/Qualifiers
1. .901
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6397884"
/clone_lib="NCI_CGAP_Co24"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: colon; Vector: pCMV-SPORT6; Site:1; NotI; Site:2; SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 247 a 246 c 225 g 180 t 3 others

ORIGIN

Query Match 25.9%; Score 443.6; DB 14; Length 901;
Best Local Similarity 71.4%; Pred. No. 1.7e-100;
Matches 640; Conservative 0; Mismatches 243; Indels 13; Gaps 4;

Qy 816 ACCCTGTTTCCCTCGGGGGAAGTCATAGTTAGGAGCACCTTCCACTTTCCTCAGG 875

Db 15 ACCAGGATTGGATTTCCTCCGAGGTTCTAGGATGAGGAGCGCTACATCTTCCCTCGG 74

Qy 876 AGAAGACACACACACCATGACCTGAGCTGGGGAGCCCAACTCCAGGCGCTCCAGCCCA 935

Db 75 ATGAGCCATACAAACCAATACCCGTGAATACGAGGACTCAAAACCTCCACTCTCCAGCCCA 134

Qy 936 AACCTGCCAGCCAGCCCTGAAATGCAAGTCTTGTACAGTTTGAAGCTAGGAACCCAC 995

Db 135 GCCCTGGCAGAGCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCAAGGATGCC 194

Qy 996 GGGAACTGACTGTGTGTCAGGGAGAGAGCTGGAGTTCTGGACCCACACAGCGGTGT 1055

Db 195 AGGAACTGACTGTGTCAGGGGAGATCTTGGAGTTCTGGATCAGACAGCGATGT 254

Qy 1056 GGCTGTTGAAGAAATGAGCGGAGCGGCTACATTCCTCAAGCAACATCTCTGGAGCC 1115

Db 255 GGCTGTTGAAGAAATGAGCGGAGCGGCTACATTCCTCAAGCAACATCTCTGGAGCC 314

Qy 1116 TACAGCCGGGGACCCCTGGGACCCAGGGCAGTCAACCTCTCTGGGTTCCTTTCGAC 1175

Db 315 TGCCAGCTGGAGCCCGGAGAGGCCACCGACAACCATCTTTAGGGCCCCCAATGCTTCG 374

Qy 1176 TTAGCTCGAGGCTTGAAGAGGTCACAGAGTGGCTGCAGGCAGAGAACTTCTCCACTGCCA 1235

Db 375 TGAGCTCAAGCTTGAGGAGGTCACAGCTGGCTTCAAGCAGAGAACTTCTCCACTGTCA 434

Qy 1236 CGGTGAGGACACTTGGTCCCTGACGGGAGCGAGCTACTTTCGATAAGACCTTGGGAGC 1295

Db 435 CTGTAGGACCTCGGGTCCCTTAATGGGAGCCAGCTACTTTCATGAGACCTGGGAGC 494

Qy 1296 TACAGATGCTATGTCACAGAGGCGCCACGAAATCTGTCCGGCTGGAGGCTGTGAGAA 1355

Db 495 TGCAGATGCTGTGTCACAGAGGCTCCACGGATCCAGGCCCGCTGGAGCTGTGAGAA 554

Qy 1356 GGATCGTGGGATAGCCCTTGGGACCCAGCTTACAGACCTTCAAGAACCCAGG-CCCCGC 1414

Db 555 GGATCGTGGGATGACTCATTTAGAGACCCAGCTCGAATACCTTAAGACCAAGGACCTCTC 614

Qy 1415 TGATCAAGATGGCAGATCTGTATACCA-TTAGAGCCCCGAGAAATTCCTTCTTGGATCC 1473

Db 615 ATGTCAAGATGGAGTATTTGGTACCCAGTTAGAGCCCCCTGGAAATCTCTT-----T 667

Qy 1474 CAGTTTGCAGCAAAACCCACACCCAGCTCACACAGCAAAACAAATGGAGCCCGCAGAG 1533

Db 668 CAGTTAGTGGCAAGCCACATATCTTGGATCACAAAGCANAGAAGATGAACAGTCCAGAG 727

Qy 1534 GCTGAAGCAACAGTGTCCCTTCTGCTGTGTGGAGCTCCCGAGTAACCACTATTATTA 1593

Db 728 ACTGAGGCAAAATGGTACCTCATTTGGG---GTTGGAGACCTCACCGGGGTACTATTATTA 783

Qy 1594 TTTTACCTCTTTCCCAACCTCGAGCATTTATGCCCTTAGCTTGTCAAGAAATCTGTTCACT 1653

Db 784 TTACCCCAACAACCTTAAGTGGCATATTATTTCCAGGCTACTCCGAGTCTCTTTAGT 843

Qy 1654 CCCTCTCTCTTCAATAAAGCATCTTCAAGCTTGTAAAGAAAAAATAANGATAA 1709

Db 844 CCAATCTCTGGCAATAAATAAACAATCGGTTTAAGTGTCTTAACAAAAAATAAATAA 899

Search completed: February 25, 2003, 16:09:44
Job time : 1702 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 17:04:28 ; Search time 30.5 Seconds

(without alignments)
10779.663 Million cell updates/sec

Title: US-09-762-021A-1

Perfect score: 3127

Sequence: 1 ggcagagcactgaagacca.....aaaaaaaaaataaataaaa 1710

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US09762021/runat_24022003_153038_7573/app_query.fasta_1.1863
-DB=PIR73 -QFMT=fastn -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021@cgn.1.1.29 @runat_24022003_153038_7573 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR73:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	558.5	17.9	821	2	S39983
2	550	17.6	822	2	138728
3	356	11.4	929	2	T27237
4	352	11.3	732	2	T27238
5	218	7.0	1464	1	S59856
6	200.5	6.4	1466	1	CGHU7L
7	192.5	6.2	847	2	P96531
8	192	6.1	660	1	Q0BE3
9	191	6.1	1300	2	T03166
10	190.5	6.1	925	2	T19361
11	190	6.1	2944	2	A54849
12	189	6.0	1188	2	S49915
13	186.5	6.0	2715	2	T13049
14	185.5	5.9	1329	2	T29074

15	185	5.9	963	2	T19140
16	185	5.9	1215	2	T32734
17	184.5	5.9	1049	1	CGBO7S
18	183.5	5.9	2774	1	A43359
19	181	5.8	1042	1	CGCH1S
20	180	5.8	1453	2	S21626
21	179	5.7	1006	2	T42731
22	175.5	5.6	1464	1	CGHU1S
23	172.5	5.5	1184	2	S50832
24	172.5	5.5	1870	2	S37671
25	172	5.5	383	2	S32975
26	172	5.5	707	2	A46302
27	171.5	5.5	862	2	T46289
28	170.5	5.5	1492	2	A40333
29	170	5.5	903	2	T00705
30	170	5.4	920	2	A45748
31	170	5.4	1872	2	S36152
32	170	5.4	3530	2	A59266
33	169.5	5.4	2649	2	T51023
34	168.5	5.4	487	2	S42442
35	168.5	5.4	990	2	T14756
36	168.5	5.4	2142	2	B35098
37	168	5.4	569	2	S28886
38	167.5	5.4	1414	1	S23809
39	167.5	5.4	1419	2	A41182
40	167.5	5.4	1487	2	B41182
41	167	5.3	924	2	S27923
42	167	5.3	1486	1	B40333
43	166	5.3	830	2	S15720
44	166	5.3	1487	1	CGHU6C
45	165.5	5.3	1880	2	T18531

ALIGNMENTS

RESULT 1

S39983

eps8 protein - mouse

C:Species: Mus musculus (house mouse)

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Nov-1999

C:Accession: S39983

R:Fazio, F.; Minichiello, L.; Matoska, V.; Castagnino, P.; Miki, T.; Wong, W.T.; di

EMBO J. 12, 3799-3808, 1993

A:Title: Eps8, a substrate for the epidermal growth factor receptor kinase, enhances

A:Reference number: S39983; MUID:94008987; PMID:8404850

A:Accession: S39983

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-821 <FAZ>

A:Cross-references: EMBL:L21671; NID:g309216; PIDN:AAA16358.1; PID:g309217

C:Superfamily: SH3 homology

F:537-584/Domain: SH3 homology <SH3>

Alignment Scores:

Pred. No.: 1.85e-31 Length: 821
Score: 558.50 Matches: 158
Percent Similarity: 38.82% Conservative: 78
Best Local Similarity: 25.99% Mismatches: 171
Query Match: 17.86% Indels: 201
DB: 2 Gaps: 17

US-09-762-021A-1 (1-1710) x S39983 (1-821)

Qy	101	GGGCTGCTATGGAAAGCCCTCCCTATGGAGGAGGACCGCTATCTGGAGCCGGGGATC	160
Db	190	GlyLysGlnLysArgArgProGluAlaArgMetIleAlaLysAlaAspProGlyIle	209
Qy	161	CCTCCAGAACAGCCACCAGGAGCCCTCCACCATCCCAAGGCC	220
Db	210	ProProPro-----ProArgAlaProAlaPro	218
Qy	221	CTGCCAGCCACACC-----AGTCC	241

Db 219 ValProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla 238
 QY 242 CGAGAACCAAGTGCCTTTACTCTGCTCCCAAGCGGTCTCTTCCCGGAGGAC--- 298
 Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGlnGlnGluThr 258
 QY 299 CCAGAG- - - - - AGGACGAGGAAGTGTCTGAACCATGTCTCTAAGG 337
 Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
 QY 338 GACATTGAGCTTTCATGGGAAGCTGGAGAGGCCAG- - - - - GCA 379
 Db 279 AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
 QY 380 AGACACGACGAGCAAGAAATTTGGGAAATAAACAAGGACCGAGGCTCTCACC--- 436
 Db 299 LysArgLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
 QY 437 - - - - - CAGGCACAGTACATTGCTTCCAGAAAGATCAAGTACAGC 478
 Db 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
 QY 479 TTCACCTCTGGGAGGCTGGCCACCTGGCTGAAGGACAAAGTGCCTCCCTGAGCTGTA 538
 Db 339 PheAsnLeuLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
 QY 539 CACATCTCTTCAAGTCCCTGAACCTTCATCTGCGCAGGTGCGCTCAGGCTGGCTAGCA 598
 Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
 QY 599 GCCCAAGTGTATCATCCCTCTCCACCCCTAAAGCTATCAACCTGTCTGCTCTA 658
 Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
 QY 659 AGCCCACTGAGAGTAACTTTGGATGGGTGGGCCCGCCAGCTGGACCACTACCGCGGCC 718
 Db 399 ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
 QY 719 GACTGGACAGGGGATGAGCCCTG---CCCTACCAACCCACATCTCAGATCAGTGGCAA 775
 Db 419 GluTrpProLysGluGlnPheIleProTyrValProArgPheArgAsnGlyTrpGlu 438
 QY 776 CTTCCA- - - - - GAGCCCTCCAGCCCAAGCACCCCTTAGGATACCGAGAC 817
 Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet- - - - - TyrGlnLeu 456
 QY 818 CCTGTTTCCCTTCGGCGGGAGTCAAT---AGGTTAGGAGGACCACTCAGCTTCTCTCAG 874
 Db 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
 QY 875 GAGAAGACACACAAACCATGACCTCAGCTCGGGAGC- - - - - 910
 Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerMet 496
 QY 911 - - - - - TCCAGGCGCTCCACGCCCAACCTGCCAGCCCTCGAAATGCAAA 964
 Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
 QY 914 AAC- - - - - TCCAGGCGCTCCACGCCCAACCTGCCAGCCCTCGAAATGCAAA 964
 Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
 QY 965 GTCTTGTACGAGTTTGAAGCTAGGAACCCAGCGGAACCTGCTGTGTCCAGGAGAGAAG 1024
 Db 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspAspVal 556
 QY 1025 CTGAGGTTCTGGACCAACGAGCGGTGTGGTGGTGAAGAAATGAGCGGACGGAGC 1084
 Db 557 LeuGluIleLeuAspAspArgArgGlnTrpTrpLysValArgAsnAlaSerGlyAspSer 576
 QY 1085 GGCTACATTCACAGCAACATCTCTGGAG- - - - - 1111
 Db 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596

QY 1111 - - - - - 1111
 Db 597 AlaAspProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
 QY 1112 - - - - - CCCCCTA 1117
 Db 617 AlaAspThrProSerAlaProSerProProProThrProAlaProValProValProLeu 636
 QY 1118 CAGCGGGGACCCCTGGG- - - - - 1135
 Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
 QY 1135 - - - - - 1135
 Db 657 AsnSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
 QY 1135 - - - - - 1135
 Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
 QY 1136 - - - - - ACCCAGGCGCAGTCACTCCCTCTCGG- - - - - GTTCCA 1165
 Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
 QY 1166 ATGCTTTCGACTTAGC- - - - - TCGAGGCTGAAGAGGTGCACAGACTGGCTGCAGCAGAG 1219
 Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
 QY 1220 AACTTCTCCACTGCCCGTGCAGGACACTTGGTCCCTGCAGCGGAGCCAGCTACTTCCG 1279
 Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
 QY 1280 ATACACCTCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCACCAATCTGTCCCGG 1339
 Db 757 LeuAsnLysAspGluLeuArgSerValCysPro- - - - - GluGlyAlaArgValPheAsnGln 775
 QY 1340 CTGAGGCTGTCTCAGAGGATGCTG 1363
 Db 776 IleThrValGlnLysAlaAlaLeu 783

RESULT 2

138728

epidermal growth factor receptor kinase substrate - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jun-2000

C:Accession: I38728

R:Wong, W.T.; Carlomagno, F.; Druck, T.; Barletta, C.; Croce, C.M.; Huebner, K.; Krau

Oncogene 9, 3057-3061, 1994.

A:Title: Evolutionary conservation of the EPS8 gene and its mapping to human chromoso

A:Reference number: I38728; MUID:94366758; PMID:8084614

A:Accession: I38728

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-822 <RES>

A:Cross-references: EMBL:U12535; NID:G530822; PID:G530823

C:Genetics:

A:Gene: Eps8

C:Superfamily: SH3 homology

Alignment Scores:

Pred. No.: 7,33e-31 Length: 822

Score: 550.00 Matches: 158

Percent Similarity: 38.71% Conservative: 77

Best Local Similarity: 26.03% Mismatches: 174

Query Match: 17.59% Indels: 198

DB: 2 Gaps: 15

US-09-762-021A-1 (1-1710) x I38728 (1-822)

QY 101 GGCCTGTCTATGAAAGCGCGCTCCCTATGTAGCAGGACGCTATCTGAGCGCGGATC 160

Db 190 GlyLysGlnLysArgArgProAspAlaLeuArgMetIleSerAsnAlaAspProSerIle 209


```

Pred. No.: 6.21e-17 Length: 732
Score: 352.00 Matches: 134
Percent Similarity: 40.27% Conservative: 75
Best Local Similarity: 25.82% Mismatches: 186
Query Match: 11.26% Indels: 124
DB: 2 Gaps: 20

US-09-762-021a-1 (1-1710) x T27238 (1-732)

Qy 224 CCAGCCACACAGTGCAGGACCAAGTGCCTTACTCTGCCTCCTCCACAGGGGTCC 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 243 ProThrHisLeuGlnAlaGlnGlnMetProPheTyrProProAspAlaSer 262
Qy 284 TCTTCCCCCGAGGCCA-----GAGAGGACGAGGAAGTCTGAACCATGTCTCTA 334
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 263 IleSerSerGluThrSerGluMetPheGluArgAspValAsnThrLeuAsnArgCysPhe 282
Qy 335 AGGACATGTAGCTGTCATGGGAAAGCTGGAGAGGCC----- 373
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 AspAspIleGluArgPheValAlaArgIleGlnSerAlaLeuAlaGlnArgGluIle 302
Qy 374 -----CAGGCAAGACACGACGAGGAAGAAGAAATTTGGCAAAAAAAC 415
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 GluGlnGlnAsnHisArgTyrArgThrAlaAsnArgArgAspLysLysAsnGlnPro 322
Qy 416 AAGGACACGAGGAGTCTC-----ACCGAGGCACAGTACATT 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 ProAspProAsnGlyIleLeuPheMetArgAlaGlnLeuProLeuGluSerGluPheVal 342
Qy 452 GACTGCTTCCAGAGATCAAGTACAGCTCAACCTCCTGGGAAGCTGGCCACCTGGCTG 511
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 AspIleLeuLysLysPheLysLeuSerPheAsnLeuLeuAlaLysLeuLysAsnHisIle 362
Qy 512 AAGGACACAAGTGCCTGAGCTCGTACACATCTCTTCAAGTCCCTCAACTTCATCTCG 571
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 363 HisGluProAsnAlaProGluLeuLeuHisPheLeuPheThrProLeuAlaValIleLeu 382
Qy 572 GCCAGTGCCTC-----GAGGCTGGCTAGCAGCCCAAGTATCTACCCCTCCTC 622
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 383 GluAlaCysHisTrpGlyLeuGlyArgAsnValAlaProThrValAlaSerProLeuLeu 402
Qy 623 ACCCTTAAGTATCAACCTGCTACAGTCTGTCTAAGCCCACTGAGAGTAACCTTTGG 682
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 403 SerLeuGluAlaArgGluLeuMetGlnAsnCysLeuThrSerHisGluSerAspIleTrp 422
Qy 683 ATGGGTTGGCCCGCTGACCACT-----ACCGCGCGGCGACTGGACAGGCAT 733
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 423 MetSerLeuGlyGluAlaTrpArgThrProProAspGluLysArgArgMetHisAlaGlu 442
Qy 734 GAGCCCTGCCCCTACCAACCCACATCTCTCAGATGACTGGCAACTTCCAGAGCCCTCC 790
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 443 LysAspLeuIleThrLysGluThrGln-----ProValProProAlaAla 459
Qy 791 -----ACCGAAGCACCTTA-----GGATACGAGGACCTGTTTCCCTTCGGCGGGA 838
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 460 ValValThrHisGlnProIleThrLysArgTyrAspProIleSerIle----- 476
Qy 839 AGTCATAGTTAGGAGACCACTTCTTCAGGAGAGACACACAACCATGACCT 898
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 477 -----SerProProGlnArgAsnAntySerHisValLys 489
Qy 899 CAGCCTGGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCACCGACCGCTG--- 955
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 490 ValThrValAsp-----SerAspThrSerProArgGlnGlnAlaPheIle 504
Qy 956 -----AAATGCAAGTCTTG-----TACGAGTTTGAAGCTAGG 988
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 505 AspAspIleValAlaLysGlyGlyLysLeuAlaValThrTyrAspArgGlyGlyGln 524
Qy 989 AACCCCGGGAACCTGCTGTCAGGAGAGAGAGTGGAGGTTCTGGACCAACGACGAG 1048
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 525 AsnThrLysGluLeuThrValHisLysGlyGlyLeuGluValIlePheAspGluArg 544

```

```

Qy 1049 CGGTGGTGGCTGGTGAAGAATGAGCGGGACGAGCGGCTACATTTCCAAAGCAATCCTG 1108
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 545 AsnTrpTrpGluCysLysAsnMetHisGlnArgValGlyTyrValProHisThrIleLeu 564
Qy 1109 GAGCCCTACAGCGGGGACCCCTGGGACCCAGGCGGCTACCCCTCTCGGGTTCCAATG 1168
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 SerMetVal-----ProPheGluGlnGlnGlnTyrAlaGlnGlnTyrAsnVal 580
Qy 1169 CTTGCACTTAGCTCGAGGCTGAAGAGCTCACAGCTGGCTGCAGGACAGA---GAACTT- 1224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 MetSerSerGlnHisAsnValIleGlnIleAsnAsp--GlyValGlyLysMetGluValP 600
Qy 1225 -----CTCCACTGCCACGGTG 1240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 600 heIleSerLysThrCysAsnPheIleLeuIleThrThrCysAsnLeuPheCysSerThr 620
Qy 1241 AGGACACTTGGTCCCTGAC----- 1260
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 620 hrLysThrAlaValProAspIleTyrArgAlaPheValMetAsnAspGluThrArgGlnM 640
Qy 1261 -----GGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACAG 1315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 640 etSerGlyThrProCysGluAsnSerArgSerPheLeuSerSerAsnThrLeuSerSer- 659
Qy 1316 GAGCCCCCAAGTCTCTCCCGCTGGAGGCTGTCTAGAAGGATGCTGGGGATAAGCCCT 1375
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 660 --SerProGluAsn-----AlaProGluIleProL 669
Qy 1376 TAGCACCCAGCTAGACACCTCCAAAGAACACAGGCCCTGATGCAAGATGGCAGATCTG 1435
   || ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 669 euLeuGluAsnLeuLysPheAsnHisAsnLysLeuTyrLeuPheHisIleLeuSerProA 689
Qy 1436 ATACCCATAGAGCCCGAGAAATCTCTCTCT----- 1467
   ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 689 snHisHis-----AsnAsnSerSerSerThrGlyGlyTyrAsnAsnGlyHisHisG 706
Qy 1468 --GGATCCCACTTTGCAGCAACCCACACCCAGCTCACACAGCAAAACAA 1518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 InGlyProGlyMetLeuProGluAspAlaProSerTyrValLysGluArgGln 723

RESULT 5
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 10-Apr-1996 #sequence.Revision 19-Apr-1996 #text_change 13-Aug-1999
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete l
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Cross-references: EMBL:X52046
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOA>
A:Cross-references: EMBL:X52046; NID:g575321; PIDN:CAA36279.1; PID:g575322
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNA
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; NID:g50476; PIDN:CAA41048.1; PID:g50477
C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3;
58/3; 673/3; 706/3; 742/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 97

```

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-154/Domain: propeptide #status predicted <PRO>
 F:32-92/Domain: von Willebrand factor type C repeat homology <WVC>
 F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Alignment Scores:
 Pred. No.: 1,77e-07 Length: 1464
 Score: 218.00 Matches: 160
 Percent Similarity: 31.77% Conservative: 37
 Best Local Similarity: 25.81% Mismatches: 220
 Query Match: 6.97% Indels: 203
 DB: 2 Gaps: 36

US-09-762-021a-1 (1-1710) x S59856 (1-1464)

QY 10 ACTGAAGACCCCTGCAGAA---GGCTCTGGAGGAAGCTGGAGCAAAAGACCTCGACT 66
 DB 645 ThrGlyGlyProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluVal 664
 QY 67 TGGAGCCCTTCAGCCAGCCAGAGATGGAGGGCCCTGCTATGGAAGCCCTCC 126
 DB 665 GlyAlaProGlyAlaProGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro 684
 QY 127 TATGGAGCA-----GGCAGCTATCTGGAGCGGGGATCCCTCCAGAACAGCC 174
 DB 685 ProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaGlyProGly----- 701
 QY 175 CCACGAGGACCCCTAGA-----GCACAGCTCCACCATCCCAAGCCCTCCAGC 228
 DB 702 ---ProGluGlyGlyLysGlyProAlaGlyProGlyProGlyProGlyAlaSerGlySer 720
 QY 229 CCACACCACTGCCCGAGAACCAAGTGCCTTTACTTGCCTCTCCAGCGGCTCTCTC 288
 DB 721 Pro-----GlyLeuGlnGly 725
 QY 289 CCCCAGGAGCCAGAGGAGGAGGAAGTGTGTAACCATGCTTAAGGGACATTTGAGCT 348
 DB 726 MetProGlyGluArgGlyGlyProGlySer-----ProGlyProLysGlyGluLysGly 743
 QY 349 GTTCATGGGAAGCTGGA---GAAGGC-----CCAGGCAAGACCCAGCAGGAAGAAATT 402
 DB 744 GluProGlyGlyAlaGlyAlaAspGlyValProGlyLysAsp----- 757
 QY 403 TGGGAAAAAACAAGGACCA---GGGAGGTCTCACCAGGACACAGTACATTTACTGCTT 459
 DB 758 -----GlyProArgGlyProAlaGlyPro----- 765
 QY 460 CCAGAAGATCAAGTACAGCTCAACCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGAC 519
 DB 766 -----IleGlyProProGlyProAlaGlyGln-----ProGlyAsp 777
 QY 520 AAGTCCCTCTGAGCTCGTACACATCTCTTCAAGTCCCTGGAACCTCACTCCGCGAGGTG 579
 DB 778 LysGlyGluGlyGlySer-----ProGlyLeu---ProGlyIleAla 790
 QY 580 ---CCCTGAGCTGGCTAGCAGCCCAAGTGTATCTCACCCCTCTCCACCCCTAAAGCTAT 636
 DB 791 GlyProArgGlyGlyProGlyGluArgGlyGluHisGlyProGlyProAlaGlyPhe 810
 QY 637 CAACCTGTCTACAGTCTGTCTAAGCCCACTGAGATCAACCTTTGGATGGGGTTG----- 691
 DB 811 ProGlyAlaProGlyGlnAsn---GlyCluProGlyAlaLysGlyGluArgGlyAlaProGl 830
 QY 692 -----GGCCCAAGCTGGACCACTACTACCGGGCCGACTGGACAGCGCATGA 735
 DB 830 yGluLysGlyGluGlyPro-ProGlyPro---AlaGlyProThrGlySerSerGlyP 849
 QY 736 GCCCTGCCCTACCAACCCACATCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCA 795
 DB 849 roAlaGlyProGlyProGlnGly---ValLysGlyGluArgGlySerPro----- 865

QY 796 AGCACCCCTTAGGATACAGGAGCCCTGTTTCCTTCGGGGGAAGTCATAGCTTAGGGAG 855
 DB 866 ---GlyProGlyThrAlaGly---PhePro---GlyGlyArg-----GlyL 879
 QY 856 CACCTCACACTTTCCTCAGGAGAGACACACAACCATGACCTCAGCCCTGGGAGCCCA- 914
 DB 879 euPro-----GlyProProG 884
 QY 915 -----ACTCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTCAAAATGCA 963
 DB 884 lyAsnAsnGlyAsnProGlyProGlyProSer----- 895
 QY 964 AGTCTGTACGAGTTTGAAGCTAGGAACCCACGAGCACTGACTGTGTCAGGAGAGAA 1023
 DB - 896 -----GlyAla-ProGlyLysAsp---GlyProProGlyPro 906
 QY 1024 GCTGAGGTTTCGGA---CCACAGCAAGGGTGTGGTGTGAAGATGAGCGGGAGC 1080
 DB 907 AlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaGlyProLys---GlyAspAla 925
 QY 1081 GAGCGCTACATTCACAAGCAACATCTCGAGCCCTACAGCCGGGAGCCCTGGGACCCA 1140
 DB 926 GlyGlnProGlyGluLysGlyProGlyAla-----GlnGlyProProGlySerPro 943
 QY 1141 GGGCAGCTCACCTCTCGGGTTCCAAATGCTTCGACTTAGCTC----- 1182
 DB 944 GlyProLeuGlyIleAlaGlyLeuThrGlyAlaArgGlyLeuAlaGlyProGlyMet 963
 QY 1183 -----GAGCCTGAAGAGTCAAGACTCGCTGCAGGAGAGACACTT 1224
 DB 964 ProGlyProArgGlySerProGlyProGlnGlyIleLysGlyGluSerGlyLysProGly 983
 QY 1225 CTCCTACTGCCAGGAGGAGACACTTGGGTCCCT--- 1257
 DB 984 AlaSerGlyHisAsnGlyGluArgGlyProProGlyProGlnGlyLeuProGlyGlnPro 1003
 QY 1258 -----GACGGGAGCCAGCTACT 1275
 DB 1004 GlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyGlnProGlyArg 1023
 QY 1276 TGCATAGACCTGGGAGCTACAGTGTATGTCACAGGA----- 1317
 DB 1024 AspGlySerProGlyLysGlyAspArgGlyGluAsnGlySerProGlyAlaProGly 1043
 QY 1318 -----GGCCCCAGCAATCTGTCTCCCGCTGGAGGTGTGTCAGAGGATGCT 1362
 DB 1044 AlaProGlyHisProGlyProProGlyProValGlyProSerGlyLysSerGlyAspArg 1063
 QY 1363 GGGGATAAG-----CCCTTAGGCACCACTTAGACACCTTCCAAAGAACCCAGGCC 1410
 DB 1064 GlyGluThrGlyProAlaGlyProSerGlyAlaProGlyProAlaGlyAlaArgGlyAla 1083
 QY 1411 CGCTGTATGCAAGATGGCAGATCTGATACCCATTTAGAGCCCGCAGAAATCTCTTCTGA 1470
 DB 1084 ProGlyProGlnGlyProArgGlyAspLysGlyGluThrGlyGluArgGlySerAsnGly 1103
 QY 1471 TCCCAAG-----TTTTCAGCAAAACCCACACCCACCTCAGCAGCAAAAAACAA 1518
 DB 1104 IleLysGlyHisArgGlyPheProGlyAsnProGlyPro----- 1116
 QY 1519 TGGACAGCCCGCAGAGCTGAAGCAACAGATGTCCCTTCTGGCTGTCTTGGAGCCCTCCCA 1578
 DB 1117 -----ProGlySerProGlyAlaAlaGlyHisGlnGlyAlaIleGlySer----- 1131
 QY 1579 GTAACCACTATTATTATTACCTCTTCCCAAACTTGGAGCATTTTGCCTAGG 1632
 DB 1132 -----ProGlyProAlaGlyProArg 1138

RESULT 6
 CSHU7L
 collagen alpha 1(III) chain precursor - human
 N:Alternate names: procollagen alpha 1(III) chain

C:Species: Homo sapiens (man)
 C:Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
 C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399; D.J.
 R:Prockop, D.J.
 Submitted to the EMBL Data Library, February 1989
 A:Reference number: S05272
 A:Accession: S05272
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1240, 'V', 1242-1466 <PRC>
 A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
 R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
 Biochem. J. 260, 509-516, 1989
 A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human procollagen.
 A:Reference number: S04642; MUID:89350838; PMID:2764886
 A:Accession: S04642
 A:Molecule type: mRNA
 A:Residues: 1-1196 <ALA>
 A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
 A:Note: the complete sequence is not shown
 R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
 Gene 78, 255-265, 1989
 A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (COL3A1).
 A:Reference number: PE0011; MUID:89378752; PMID:2777083
 A:Accession: PE0011
 A:Molecule type: DNA
 A:Residues: 1-176 <BEN>
 A:Cross-references: GB:M26939; NID:g180813; PIDN:AA52040.1; PID:g180814
 R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
 Nucleic Acids Res. 16, 7201, 1988
 A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human procollagen type III.
 A:Reference number: S01726; MUID:88303360; PMID:3405773
 A:Accession: S01726
 A:Molecule type: mRNA
 A:Residues: 1-170 <TOM>
 A:Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
 A:Note: the authors translated the codon CAG for residue 154 as His
 R:Janeczko, R.A.; Ramirez, F.
 Nucleic Acids Res. 17, 6742, 1989
 A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
 A:Reference number: S04887; MUID:89386015; PMID:2780304
 A:Accession: S04887
 A:Molecule type: mRNA
 A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'Y', 635-636, 'G', 637-638, 'G', 639-640, 'G', 641-642, 'G', 643-644, 'G', 645-646, 'G', 647-648, 'G', 649-650, 'G', 651-652, 'G', 653-654, 'G', 655-656, 'G', 657-658, 'G', 659-660, 'G', 661-662, 'G', 663-664, 'G', 665-666, 'G', 667-668, 'G', 669-670, 'G', 671-672, 'G', 673-674, 'G', 675-676, 'G', 677-678, 'G', 679-680, 'G', 681-682, 'G', 683-684, 'G', 685-686, 'G', 687-688, 'G', 689-690, 'G', 691-692, 'G', 693-694, 'G', 695-696, 'G', 697-698, 'G', 699-700, 'G', 701-702, 'G', 703-704, 'G', 705-706, 'G', 707-708, 'G', 709-710, 'G', 711-712, 'G', 713-714, 'G', 715-716, 'G', 717-718, 'G', 719-720, 'G', 721-722, 'G', 723-724, 'G', 725-726, 'G', 727-728, 'G', 729-730, 'G', 731-732, 'G', 733-734, 'G', 735-736, 'G', 737-738, 'G', 739-740, 'G', 741-742, 'G', 743-744, 'G', 745-746, 'G', 747-748, 'G', 749-750, 'G', 751-752, 'G', 753-754, 'G', 755-756, 'G', 757-758, 'G', 759-760, 'G', 761-762, 'G', 763-764, 'G', 765-766, 'G', 767-768, 'G', 769-770, 'G', 771-772, 'G', 773-774, 'G', 775-776, 'G', 777-778, 'G', 779-780, 'G', 781-782, 'G', 783-784, 'G', 785-786, 'G', 787-788, 'G', 789-790, 'G', 791-792, 'G', 793-794, 'G', 795-796, 'G', 797-798, 'G', 799-800, 'G', 801-802, 'G', 803-804, 'G', 805-806, 'G', 807-808, 'G', 809-810, 'G', 811-812, 'G', 813-814, 'G', 815-816, 'G', 817-818, 'G', 819-820, 'G', 821-822, 'G', 823-824, 'G', 825-826, 'G', 827-828, 'G', 829-830, 'G', 831-832, 'G', 833-834, 'G', 835-836, 'G', 837-838, 'G', 839-840, 'G', 841-842, 'G', 843-844, 'G', 845-846, 'G', 847-848, 'G', 849-850, 'G', 851-852, 'G', 853-854, 'G', 855-856, 'G', 857-858, 'G', 859-860, 'G', 861-862, 'G', 863-864, 'G', 865-866, 'G', 867-868, 'G', 869-870, 'G', 871-872, 'G', 873-874, 'G', 875-876, 'G', 877-878, 'G', 879-880, 'G', 881-882, 'G', 883-884, 'G', 885-886, 'G', 887-888, 'G', 889-890, 'G', 891-892, 'G', 893-894, 'G', 895-896, 'G', 897-898, 'G', 899-900, 'G', 901-902, 'G', 903-904, 'G', 905-906, 'G', 907-908, 'G', 909-910, 'G', 911-912, 'G', 913-914, 'G', 915-916, 'G', 917-918, 'G', 919-920, 'G', 921-922, 'G', 923-924, 'G', 925-926, 'G', 927-928, 'G', 929-930, 'G', 931-932, 'G', 933-934, 'G', 935-936, 'G', 937-938, 'G', 939-940, 'G', 941-942, 'G', 943-944, 'G', 945-946, 'G', 947-948, 'G', 949-950, 'G', 951-952, 'G', 953-954, 'G', 955-956, 'G', 957-958, 'G', 959-960, 'G', 961-962, 'G', 963-964, 'G', 965-966, 'G', 967-968, 'G', 969-970, 'G', 971-972, 'G', 973-974, 'G', 975-976, 'G', 977-978, 'G', 979-980, 'G', 981-982, 'G', 983-984, 'G', 985-986, 'G', 987-988, 'G', 989-990, 'G', 991-992, 'G', 993-994, 'G', 995-996, 'G', 997-998, 'G', 999-1000, 'G', 1001-1002, 'G', 1003-1004, 'G', 1005-1006, 'G', 1007-1008, 'G', 1009-1010, 'G', 1011-1012, 'G', 1013-1014, 'G', 1015-1016, 'G', 1017-1018, 'G', 1019-1020, 'G', 1021-1022, 'G', 1023-1024, 'G', 1025-1026, 'G', 1027-1028, 'G', 1029-1030, 'G', 1031-1032, 'G', 1033-1034, 'G', 1035-1036, 'G', 1037-1038, 'G', 1039-1040, 'G', 1041-1042, 'G', 1043-1044, 'G', 1045-1046, 'G', 1047-1048, 'G', 1049-1050, 'G', 1051-1052, 'G', 1053-1054, 'G', 1055-1056, 'G', 1057-1058, 'G', 1059-1060, 'G', 1061-1062, 'G', 1063-1064, 'G', 1065-1066, 'G', 1067-1068, 'G', 1069-1070, 'G', 1071-1072, 'G', 1073-1074, 'G', 1075-1076, 'G', 1077-1078, 'G', 1079-1080, 'G', 1081-1082, 'G', 1083-1084, 'G', 1085-1086, 'G', 1087-1088, 'G', 1089-1090, 'G', 1091-1092, 'G', 1093-1094, 'G', 1095-1096, 'G', 1097-1098, 'G', 1099-1100, 'G', 1101-1102, 'G', 1103-1104, 'G', 1105-1106, 'G', 1107-1108, 'G', 1109-1110, 'G', 1111-1112, 'G', 1113-1114, 'G', 1115-1116, 'G', 1117-1118, 'G', 1119-1120, 'G', 1121-1122, 'G', 1123-1124, 'G', 1125-1126, 'G', 1127-1128, 'G', 1129-1130, 'G', 1131-1132, 'G', 1133-1134, 'G', 1135-1136, 'G', 1137-1138, 'G', 1139-1140, 'G', 1141-1142, 'G', 1143-1144, 'G', 1145-1146, 'G', 1147-1148, 'G', 1149-1150, 'G', 1151-1152, 'G', 1153-1154, 'G', 1155-1156, 'G', 1157-1158, 'G', 1159-1160, 'G', 1161-1162, 'G', 1163-1164, 'G', 1165-1166, 'G', 1167-1168, 'G', 1169-1170, 'G', 1171-1172, 'G', 1173-1174, 'G', 1175-1176, 'G', 1177-1178, 'G', 1179-1180, 'G', 1181-1182, 'G', 1183-1184, 'G', 1185-1186, 'G', 1187-1188, 'G', 1189-1190, 'G', 1191-1192, 'G', 1193-1194, 'G', 1195-1196, 'G', 1197-1198, 'G', 1199-1200, 'G', 1201-1202, 'G', 1203-1204, 'G', 1205-1206, 'G', 1207-1208, 'G', 1209-1210, 'G', 1211-1212, 'G', 1213-1214, 'G', 1215-1216, 'G', 1217-1218, 'G', 1219-1220, 'G', 1221-1222, 'G', 1223-1224, 'G', 1225-1226, 'G', 1227-1228, 'G', 1229-1230, 'G', 1231-1232, 'G', 1233-1234, 'G', 1235-1236, 'G', 1237-1238, 'G', 1239-1240, 'G', 1241-1242, 'G', 1243-1244, 'G', 1245-1246, 'G', 1247-1248, 'G', 1249-1250, 'G', 1251-1252, 'G', 1253-1254, 'G', 1255-1256, 'G', 1257-1258, 'G', 1259-1260, 'G', 1261-1262, 'G', 1263-1264, 'G', 1265-1266, 'G', 1267-1268, 'G', 1269-1270, 'G', 1271-1272, 'G', 1273-1274, 'G', 1275-1276, 'G', 1277-1278, 'G', 1279-1280, 'G', 1281-1282, 'G', 1283-1284, 'G', 1285-1286, 'G', 1287-1288, 'G', 1289-1290, 'G', 1291-1292, 'G', 1293-1294, 'G', 1295-1296, 'G', 1297-1298, 'G', 1299-1300, 'G', 1301-1302, 'G', 1303-1304, 'G', 1305-1306, 'G', 1307-1308, 'G', 1309-1310, 'G', 1311-1312, 'G', 1313-1314, 'G', 1315-1316, 'G', 1317-1318, 'G', 1319-1320, 'G', 1321-1322, 'G', 1323-1324, 'G', 1325-1326, 'G', 1327-1328, 'G', 1329-1330, 'G', 1331-1332, 'G', 1333-1334, 'G', 1335-1336, 'G', 1337-1338, 'G', 1339-1340, 'G', 1341-1342, 'G', 1343-1344, 'G', 1345-1346, 'G', 1347-1348, 'G', 1349-1350, 'G', 1351-1352, 'G', 1353-1354, 'G', 1355-1356, 'G', 1357-1358, 'G', 1359-1360, 'G', 1361-1362, 'G', 1363-1364, 'G', 1365-1366, 'G', 1367-1368, 'G', 1369-1370, 'G', 1371-1372, 'G', 1373-1374, 'G', 1375-1376, 'G', 1377-1378, 'G', 1379-1380, 'G', 1381-1382, 'G', 1383-1384, 'G', 1385-1386, 'G', 1387-1388, 'G', 1389-1390, 'G', 1391-1392, 'G', 1393-1394, 'G', 1395-1396, 'G', 1397-1398, 'G', 1399-1400, 'G', 1401-1402, 'G', 1403-1404, 'G', 1405-1406, 'G', 1407-1408, 'G', 1409-1410, 'G', 1411-1412, 'G', 1413-1414, 'G', 1415-1416, 'G', 1417-1418, 'G', 1419-1420, 'G', 1421-1422, 'G', 1423-1424, 'G', 1425-1426, 'G', 1427-1428, 'G', 1429-1430, 'G', 1431-1432, 'G', 1433-1434, 'G', 1435-1436, 'G', 1437-1438, 'G', 1439-1440, 'G', 1441-1442, 'G', 1443-1444, 'G', 1445-1446, 'G', 1447-1448, 'G', 1449-1450, 'G', 1451-1452, 'G', 1453-1454, 'G', 1455-1456, 'G', 1457-1458, 'G', 1459-1460, 'G', 1461-1462, 'G', 1463-1464, 'G', 1465-1466, 'G', 1467-1468, 'G', 1469-1470, 'G', 1471-1472, 'G', 1473-1474, 'G', 1475-1476, 'G', 1477-1478, 'G', 1479-1480, 'G', 1481-1482, 'G', 1483-1484, 'G', 1485-1486, 'G', 1487-1488, 'G', 1489-1490, 'G', 1491-1492, 'G', 1493-1494, 'G', 1495-1496, 'G', 1497-1498, 'G', 1499-1500, 'G', 1501-1502, 'G', 1503-1504, 'G', 1505-1506, 'G', 1507-1508, 'G', 1509-1510, 'G', 1511-1512, 'G', 1513-1514, 'G', 1515-1516, 'G', 1517-1518, 'G', 1519-1520, 'G', 1521-1522, 'G', 1523-1524, 'G', 1525-1526, 'G', 1527-1528, 'G', 1529-1530, 'G', 1531-1532, 'G', 1533-1534, 'G', 1535-1536, 'G', 1537-1538, 'G', 1539-1540, 'G', 1541-1542, 'G', 1543-1544, 'G', 1545-1546, 'G', 1547-1548, 'G', 1549-1550, 'G', 1551-1552, 'G', 1553-1554, 'G', 1555-1556, 'G', 1557-1558, 'G', 1559-1560, 'G', 1561-1562, 'G', 1563-1564, 'G', 1565-1566, 'G', 1567-1568, 'G', 1569-1570, 'G', 1571-1572, 'G', 1573-1574, 'G', 1575-1576, 'G', 1577-1578, 'G', 1579-1580, 'G', 1581-1582, 'G', 1583-1584, 'G', 1585-1586, 'G', 1587-1588, 'G', 1589-1590, 'G', 1591-1592, 'G', 1593-1594, 'G', 1595-1596, 'G', 1597-1598, 'G', 1599-1600, 'G', 1601-1602, 'G', 1603-1604, 'G', 1605-1606, 'G', 1607-1608, 'G', 1609-1610, 'G', 1611-1612, 'G', 1613-1614, 'G', 1615-1616, 'G', 1617-1618, 'G', 1619-1620, 'G', 1621-1622, 'G', 1623-1624, 'G', 1625-1626, 'G', 1627-1628, 'G', 1629-1630, 'G', 1631-1632, 'G', 1633-1634, 'G', 1635-1636, 'G', 1637-1638, 'G', 1639-1640, 'G', 1641-1642, 'G', 1643-1644, 'G', 1645-1646, 'G', 1647-1648, 'G', 1649-1650, 'G', 1651-1652, 'G', 1653-1654, 'G', 1655-1656, 'G', 1657-1658, 'G', 1659-1660, 'G', 1661-1662, 'G', 1663-1664, 'G', 1665-1666, 'G', 1667-1668, 'G', 1669-1670, 'G', 1671-1672, 'G', 1673-1674, 'G', 1675-1676, 'G', 1677-1678, 'G', 1679-1680, 'G', 1681-1682, 'G', 1683-1684, 'G', 1685-1686, 'G', 1687-1688, 'G', 1689-1690, 'G', 1691-1692, 'G', 1693-1694, 'G', 1695-1696, 'G', 1697-1698, 'G', 1699-1700, 'G', 1701-1702, 'G', 1703-1704, 'G', 1705-1706, 'G', 1707-1708, 'G', 1709-1710, 'G', 1711-1712, 'G', 1713-1714, 'G', 1715-1716, 'G', 1717-1718, 'G', 1719-1720, 'G', 1721-1722, 'G', 1723-1724, 'G', 1725-1726, 'G', 1727-1728, 'G', 1729-1730, 'G', 1731-1732, 'G', 1733-1734, 'G', 1735-1736, 'G', 1737-1738, 'G', 1739-1740, 'G', 1741-1742, 'G', 1743-1744, 'G', 1745-1746, 'G', 1747-1748, 'G', 1749-1750, 'G', 1751-1752, 'G', 1753-1754, 'G', 1755-1756, 'G', 1757-1758, 'G', 1759-1760, 'G', 1761-1762, 'G', 1763-1764, 'G', 1765-1766, 'G', 1767-1768, 'G', 1769-1770, 'G', 1771-1772, 'G', 1773-1774, 'G', 1775-1776, 'G', 1777-1778, 'G', 1779-1780, 'G', 1781-1782, 'G', 1783-1784, 'G', 1785-1786, 'G', 1787-1788, 'G', 1789-1790, 'G', 1791-1792, 'G', 1793-1794, 'G', 1795-1796, 'G', 1797-1798, 'G', 1799-1800, 'G', 1801-1802, 'G', 1803-1804, 'G', 1805-1806, 'G', 1807-1808, 'G', 1809-1810, 'G', 1811-1812, 'G', 1813-1814, 'G', 1815-1816, 'G', 1817-1818, 'G', 1819-1820, 'G', 1821-1822, 'G', 1823-1824, 'G', 1825-1826, 'G', 1827-1828, 'G', 1829-1830, 'G', 1831-1832, 'G', 1833-1834, 'G', 1835-1836, 'G', 1837-1838, 'G', 1839-1840, 'G', 1841-1842, 'G', 1843-1844, 'G', 1845-1846, 'G', 1847-1848, 'G', 1849-1850, 'G', 1851-1852, 'G', 1853-1854, 'G', 1855-1856, 'G', 1857-1858, 'G', 1859-1860, 'G', 1861-1862, 'G', 1863-1864, 'G', 1865-1866, 'G', 1867-1868, 'G', 1869-1870, 'G', 1871-1872, 'G', 1873-1874, 'G', 1875-1876, 'G', 1877-1878, 'G', 1879-1880, 'G', 1881-1882, 'G', 1883-1884, 'G', 1885-1886, 'G', 1887-1888, 'G', 1889-1890, 'G', 1891-1892, 'G', 1893-1894, 'G', 1895-1896, 'G', 1897-1898, 'G', 1899-1900, 'G', 1901-1902, 'G', 1903-1904, 'G', 1905-1906, 'G', 1907-1908, 'G', 1909-1910, 'G', 1911-1912, 'G', 1913-1914, 'G', 1915-1916, 'G', 1917-1918, 'G', 1919-1920, 'G', 1921-1922, 'G', 1923-1924, 'G', 1925-1926, 'G', 1927-1928, 'G', 1929-1930, 'G', 1931-1932, 'G', 1933-1934, 'G', 1935-1936, 'G', 1937-1938, 'G', 1939-1940, 'G', 1941-1942, 'G', 1943-1944, 'G', 1945-1946, 'G', 1947-1948, 'G', 1949-1950, 'G', 1951-1952, 'G', 1953-1954, 'G', 1955-1956, 'G', 1957-1958, 'G', 1959-1960, 'G', 1961-1962, 'G', 1963-1964, 'G', 1965-1966, 'G', 1967-1968, 'G', 1969-1970, 'G', 1971-1972, 'G', 1973-1974, 'G', 1975-1976, 'G', 1977-1978, 'G', 1979-1980, 'G', 1981-1982, 'G', 1983-1984, 'G', 1985-1986, 'G', 1987-1988, 'G', 1989-1990, 'G', 1991-1992, 'G', 1993-1994, 'G', 1995-1996, 'G', 1997-1998, 'G', 1999-2000, 'G', 2001-2002, 'G', 2003-2004, 'G', 2005-2006, 'G', 2007-2008, 'G', 2009-2010, 'G', 2011-2012, 'G', 2013-2014, 'G', 2015-2016, 'G', 2017-2018, 'G', 2019-2020, 'G', 2021-2022, 'G', 2023-2024, 'G', 2025-2026, 'G', 2027-2028, 'G', 2029-2030, 'G', 2031-2032, 'G', 2033-2034, 'G', 2035-2036, 'G', 2037-2038, 'G', 2039-2040, 'G', 2041-2042, 'G', 2043-2044, 'G', 2045-2046, 'G', 2047-2048, 'G', 2049-2050, 'G', 2051-2052, 'G', 2053-2054, 'G', 2055-2056, 'G', 2057-2058, 'G', 2059-2060, 'G', 2061-2062, 'G', 2063-2064, 'G', 2065-2066, 'G', 2067-2068, 'G', 2069-2070, 'G', 2071-2072, 'G', 2073-2074, 'G', 2075-2076, 'G', 2077-2078, 'G', 2079-2080, 'G', 2081-2082, 'G', 2083-2084, 'G', 2085-2086, 'G', 2087-2088, 'G', 2089-2090, 'G', 2091-2092, 'G', 2093-2094, 'G', 2095-2096, 'G', 2097-2098, 'G', 2099-2100, 'G', 2101-2102, 'G', 2103-2104, 'G', 2105-2106, 'G', 2107-2108, 'G', 2109-2110, 'G', 2111-2112, 'G', 2113-2114, 'G', 2115-2116, 'G', 2117-2118, 'G', 2119-2120, 'G', 2121-2122, 'G', 2123-2124, 'G', 2125-2126, 'G', 2127-2128, 'G', 2129-2130, 'G', 2131-2132, 'G', 2133-2134, 'G', 2135-2136, 'G', 2137-2138, 'G', 2139-2140, 'G', 2141-2142, 'G', 2143-2144, 'G', 2145-2146, 'G', 2147-2148, 'G', 2149-2150, 'G', 2151-2152, 'G', 2153-2154, 'G', 2155-2156, 'G', 2157-2158, 'G', 2159-2160, 'G', 2161-2162, 'G', 2163-2164, 'G', 2165-2166, 'G', 2167-2168, 'G', 2169-2170, 'G', 2171-2172, 'G', 2173-2174, 'G', 2175-2176, 'G', 2177-2178, 'G', 2179-2180, 'G', 2181-2182, 'G', 2183-2184, 'G', 2185-2186, 'G', 2187-2188, 'G', 2189-2190, 'G', 2191-2192, 'G', 2193-2194, 'G', 2195-2196, 'G', 2197-2198, 'G', 2199-2200, 'G', 2201-2202, 'G', 2203-2204, 'G', 2205-2206, 'G', 2207-2208, 'G', 2209-2210, 'G', 2211-2212, 'G', 2213-2214, 'G', 2215-2216, 'G', 2217-2218, 'G', 2219-2220, 'G', 2221-2222, 'G', 2223-2224, 'G', 2225-2226, 'G', 2227-2228, 'G', 2229-2230, 'G', 2231-2232, 'G', 2233-2234, 'G', 2235-2236, 'G', 2237-2238, 'G', 2239-2240, 'G', 2241-2242, 'G', 2243-2244, 'G', 2245-2246, 'G', 2247-2248, 'G', 2249-2250, 'G', 2251-2252, 'G', 2253-2254, 'G', 2255-2256, 'G', 2257-2258, 'G', 2259

A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1161-1200 <MIS>
 A:Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:g180416
 R:Emanuel, B.S.; Cannizzaro, L.A.; Sever, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A:Reference number: 159025; MUID:85216505; PMID:3858826
 A:Accession: I79359
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1165-1196 <MA>
 A:Cross-references: GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:g180418
 R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
 J. Biol. Chem. 260, 4357-4363, 1985
 A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. B
 A:Reference number: A92516; MUID:85157600; PMID:2579949
 A:Accession: A92516
 A:Molecule type: DNA
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
 A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
 A:Experimental source: liver
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
 ation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 3-hydroxylated. About 15% of the lysines are 5-hydroxylated and some are subsequently
 C:Genetics:
 A:Gene: GDB:COL3A1
 A:Cross-references: GDB:118729; OMIM:120180
 A:Map position: 2q31-2q31
 A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
 er of their length, is formed with desmosine cross-links made from lysine and allysine r
 C:Function:
 A:Description: structural component of extracellular fibrous polymer that maintains inte
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
 F:31-91/Domain: von Willebrand factor type C repeat homology <WV>
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:154-167/Region: amino-terminal nonhelical telopeptide
 F:168-1196/Region: helical
 F:1091-1093/Region: cell attachment (R-G-D) motif
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
 F:1223-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:153-154/Cleavage site: pro-Gln (procollagen N-endopeptidase) #status predicted
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
 F:161,1212/Modified site: allysine (Lys) #status predicted
 F:263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:263/Binding site: carbohydurate (Lys) (covalent) #status experimental
 F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
 F:1106/Binding site: carbohydurate (Lys) (covalent) #status predicted

Alignment Scores:

Score:	3.01e-06	Length:	1466
Seq. No.:	200.50	Matches:	157
Percent Similarity:	32.01%	Conservative:	37
Best Local Similarity:	25.91%	Mismatches:	224
Query Match:	6.41%	Indels:	188
DB:	1	Gaps:	34

US-09-762-021A-1 (1-1710) x CGH07L (1-1466)

QY 10 ACTGAAGACGACGCTGCAGAA---GCCTCTGGAGGAGAGCTGGAGCAAGACCTCGACT 66

Db 646 ThrGlyGlyProGlyGluAsnGlyLysProGlyGluProGlyGlyProGlyProGlyLysGlyAspAla 665

QY 67 TGGAGCCCTTACGCCAGCCAGGACAGATGGAGGGGGCCTGTCTATGGAAGCCGCTCCC 126

Db	666	GlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGlyGluArgGlyPro	685
QY	127	TATGGA-----CGAGCGACGCTATCTGGAGCGGGGATCCCTCCAGAACGCC	174
Db	686	ProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProGly-----	702
QY	175	CCACCAGAGGACCCCTAGA-----GCACAGCCTCCACCACCTCCCAAGGCCCTCCGACG	228
Db	703	---ProGluGlyGlyLysGlyAlaAlaGlyProGlyProGlyAlaAlaGlyThr	721
QY	229	CCACCAGCAGTCCCGGAGAACCAAGTCCTTACTCTGCCTCTCTCAAGGGGTCCTCTTC	288
Db	722	Pro-----GlyLeuGlnGly	726
QY	289	CCCCGAGGACCCAGAGAGGAGGAGCACTGCTGAACCATGCTCCTAAGGACATTCAGCT	348
Db	727	MetProGlyGluArgGlyGlyLeuGlySer-----ProGlyProLysGlyAsp-----	742
QY	349	GTTTCATGGAAAGCTGGAGAGGCCCGCCAGCAAGACCAGCAGCAGAGAGAGAAATTTGGGA	408
Db	743	---LysGlyGluProGly---GlyProGlyAlaAspGlyVal-----Pro	755
QY	409	AAAAAACAGAGGACCA---GGGAGGTCTCACCCAGGACACAGTACATTGACTGCTTCAGAA	465
Db	756	GlyLysAspGlyProArgGlyProThrGlyPro-----AlaProGlyLeu---ProGlyIleAlaGlyPro	793
QY	466	GATCAAGTACAGCTTCAACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGACACAAAGTCG	525
Db	767	-----IleGlyProProGlyProAlaGlyGln-----ProGlyAspLysGly	780
QY	526	CCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAATCTTCCTGGCCAGGTG---CCC	582
Db	781	GluGlyGly-----AlaProGlyLeu---ProGlyIleAlaGlyPro	793
QY	583	TGAGCTGCGCTAGCAGCCCAAGTATCTACCCCTCTCACCCTTAAGCTATATCAACCT	642
Db	794	ArgGlySerProGlyGluArgGlyGlyThrGlyProProGlyProAlaGlyPheProGly	813
QY	643	GCTACAGCTCTGCTTAAGCCCACTCCAGAGTAGTAACCTTTGGATGGGGTTG-----	691
Db	814	AlaProGlyGlnAsn---GlyGluProGlyGlyLysGlyGlyAlaProGlyGlyLys	833
QY	692	-----GGCCAGCCTGGACCACTAGCCGGCGGCGCTGCACAGCGCATGAGCCCT	741
Db	833	sglyGluGlyGlyPro---ProGly---ValaGlyProProGlyGlySerGlyProAlaG	852
QY	742	GCCCTACCAACCCATCTCAGATGACTGGCAACTTCCAGAGCCCT-----	788
Db	852	lyProProGlyProGlnGly---ValLysGlyGluArgGlySerProGlyGlyProGlyA	871
QY	789	-----CCAGCCCAAGCACCCCTTAGGATACCA	813
Db	871	laAlaGlyPheProGlyAlaArgGlyLeuProGlyPro---ProGlySerAsnGlyAsnPro	890
QY	814	GGACCTGTTTCCCTTCGGCGGGGAGTATAGTTTAGGGAGCACCCTCACACTTTCCTCA	873
Db	891	GlyProProGlyProSerGly-----SerPro	899
QY	874	GGAGAGACACACACCATGACCTCA-----GCC	903
Db	900	GlyLysAspGlyProProGlyProAlaGlyAsnThrGlyAlaProGlySerProGlyVal	919
QY	904	TGGGACCCCACTCCAGGCCCTCCAGCCC-----CAAAACC	939
Db	920	SerGlyProLysGlyAspAlaGlyGlnProGlyGlyLysGlySerProGlyAlaGlnGly	939
QY	940	TGCCCGACCCGCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACCGGA	999
Db	940	ProProGlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAla	959
QY	1000	ACTGACTGTGTCCAGGGAGAGAGCTGGAGGTTCTTGGACCACAGCAAGCGGTGGTGGT	1059

```

Db 960 GlyProProGlyMetProGlyProArgGlySerProGlyProGlnGly-----Vallys 977
Qy 1060 GGTGAAGAAATGAGGC-----GGCAGGAGCGGCTACATCTCCAAGCAACATCTCGGAGCC 1113
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 978 GlyGluSer---GlyLysProGlyAlaAsnGlyLeuSerGlyGluArgGlyProGly 996
Qy 1114 CCT-----ACAGCCGGGAGCC-----TGGGACCCA 1140
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 997 ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGluProGlyArgAspGlyAsnPro 1016
Qy 1141 GGGCCAGTCACCCCTCTCGGTTCCTCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTAC 1200
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1017 GlySerAspGlyLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGlu 1036
Qy 1201 AGACTGGCTGAGGAGCAGAGAATCTCTCCACTGCCACGGTGAGGACACTTGGGTCCCTGAC 1260
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1037 -----Asn 1037
Qy 1261 GGGAGGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATCTCCACAGGAGC 1320
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1038 GlySerPro-----GlyAlaProGlyAlaProGlyHisProGly 1050
Qy 1321 CCCAGCAATCTGTCGGCTGGAGGCTGTCAGAAGATGCTGGGGATAG----- 1371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1051 ProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluSerGlyProAla 1070
Qy 1372 ---CCCTTAGCACCAGCTTAGACACCTCCAAGAACCCAGCCGCTGATGCAAGATGGC 1428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1071 GlyProAlaGlyAlaProGlyProAlaGlySerArgGlyAlaProGlyProGlnGlyPro 1090
Qy 1429 AGATCTGATACCCATTAGACCCCGAGAAATTCCTTCTGCTGACCCAG----- 1476
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1091 ArgGlyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyLeLysGlyHisArgGly 1110
Qy 1477 TTTCGCAACACCC-----ACACCCAGCTCACACAGCAAAACATGGACA 1524
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1111 PheProGlyAsnProGlyAlaProGlySerPro----- 1121
Qy 1525 GCCCAGAGGCTGAAGCAACAGTGTC-----CCTCTGCGCTGT 1563
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1122 GlyProAlaGlyGlnGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyPro 1141
Qy 1564 GTTGGAGCCTCC 1575
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1142 ValGlyProSer 1145

RESULT 7
F96531
hypothetical protein F13F21.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96531
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F96531
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-847 <SPO>
A:Cross-references: GB:AE005173; NID:g5430752; PIDN:AAD43152.1; GSPDB:GN00141
C:Genetics:
A:Gene: F13F21.7
A:Map position: 1
Alignment Scores:

```

```

Pred. No.: 1.05e-05 Length: 847
Score: 192.50 Matches: 134
Percent Similarity: 34.29% Conservative: 58
Best Local Similarity: 23.93% Mismatches: 210
Query Match: 6.16% Indels: 159
DB: 2 Gaps: 25

US-09-762-021A-1 (1-1710) x F96531 (1-847)

Qy 70 AGGCCTTCAGCCAGCCAGGACAGATG----- 96
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 360 ArgProSerGlnLysProAlaLysGlnCysLeuProValSerArgProValAspCys 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 97 ---GAGGGGCGCTCTATGGAAGGCGCTCCCTATGGAGCAGCAGCGCTATCTGGAGCC 153
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 380 SerLysAspLysCysSerGlyGlySer-----AsnGlyGlySerSerProSer 395
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 154 GGGATCCCTCCAGACAGCCACAGAGGACCCCTAGACACAGCCT----- 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 396 ProAsnProProArgThrSerGluProLysProSerLysProGluProValMetProLys 415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 202 CCCACCATCCCAAGGCCCTGCCCACGCCACAGCAGTCCCGAGAACCAAGTGCCTTAC 261
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 416 ProSerAspSerSerLysProGluThrPro-LysThrProGluGlnProSerProLysPr 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 262 TCTGCCCTCTCCAAAGGCGCTCTTCCCGGAGGACCCAGAGGAGGACGAGGAAGTCT 321
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 435 OGlnProProLysHisGluSerProLysProGluGluProGluAsnLysHisGluLeuPr 455
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 322 -----GAACCATGCTCTAAGGACATTTAGCTGTTTCATGGGAAAGCTGGAGA 368
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 455 oLysGlnLysGluSerProLysProGln-----ProSerLysProGluAs 470
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 369 AGCCCCAGGCAAGACCCAGCAGGAGGAAGAAATTTGGGAAAAAACAAGCAGGAG 428
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 470 pSerProLysProGluGlnProLysProGluGluSerProLysProGluGlnPro----- 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 429 GTCTCACCAGGACAGTACATTTGACTGCTTCCAGAGAGTCAAGTACAGCTTCAACCTCC 488
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 489 -----GlnIleProGluProThrLysProValSerProPr 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 489 TGGGAAGCTGGCCACCTGGCTGAAGGAGAC----- 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 500 oAsnGluAlaGlnGlyProThrProAspAspProTyAspAlaSerProValLysAsnAr 520
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 520 ---AAGTGGCCCTGAGCTCGTACACATCTCTTCAAGTCCCTCAACTTATCTTGCCCA 575
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 520 gArgSerProProProLysValGluAspThrArgValPro-----ProProGl 537
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 576 GGTGCCCTGAGGCTGGGCTAGCAGCCCAAGTGATCTC---ACCCTCTCTCACCCCTAAAG 632
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 537 nProProMetProSerProSerProProProSerProLysSerProProProProValH 557
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 633 CTATCAACCTGCTACAGTCTCTGCTTAAGCCACCTGAGAGTAACCTTTGGATGGGTGG 692
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 557 sSerProProProValTySer-SerProProProHisValTy-----S 574
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 693 GCCCAGCTGGACCACTAGCGGGCCCACTGGACAGCGGATGAGCCCTGCCCTACCAAC 752
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 574 erPro-ProProProValAlaSerProPro-----ProProSerProProPro 589
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 753 CCA---CATTCTCTAGAGTCTGCGCACTTCAGAGCCCTCCAGCAGCCACCCCTTAGAT 809
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 590 ProValHisSerProProProProProProValPheSerProPro-ProProValPheSerPr 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 810 ACCAGGACCTGTTTCCCTTCGGGGGGAAGTATAGTTAGGAGACCTTCACACTTTC 869
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 609 oPro-ProPro-----SerProValTySerPr 618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 870 CTAGGAGAGACACACAACCATGACCTCAGCTGGGGGAGCCCACTCCAGGCCCTCCA 929
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 618 roProProSerHisSerProProProProValTySerProProProProProThrPheS 638
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

Alignment Scores:

QY 930 GCCCAACCTGCC-----CAGCCAGCCCTGAAATGCAAGTCTTGTACGAGT 977
Db 638 erProProThrHisAsnThrAsnGlnProProMet 650
QY 978 TTGAAGCTAGGAACCCACGGGAACCTGACTGTGTCCAGGGAGAGAGCTGGAGGTTCTGG 1037
Db 651 --GlyAlaProThrProThrGlnAlaProThrProSerSerGluThrThrGlnValProT 670
QY 1038 ACCACAGCAAGCGGTGGTGGTGTGAAGAATGAGCGGGAGCGGAGCGGTACATTCCAA 1097
Db 670 hrProSer-----SerGluSerAspGln-----S 678
QY 1098 GCAACATCTGTGAGCCCTACAGCGGGGAGCCCTGGGACCCAGCGGCGAGTCACTCTC 1157
Db 678 erGlnIleLeuSerProValGlnAlaProThrPro---ValGlnSerSerThrPro--- 695
QY 1158 GGGTTCCAATGCTTCGACTTACCTGAGGCGCTGAAGAGTC-----ACAG 1202
Db 696 -----SerSerGluProThrGlnValProThrProSerSerSerG 709
QY 1203 ACTGGCTGCAGGACAGAACTTCTCCACTGCCACGGTGAGGACACTTGGTCCCTGACGG 1262
Db 709 luSerTyrGlnAlaProAsnLeuSerProValGlnAlaProThr----- 723
QY 1263 GGAGCCAGCTACTTCGCATAAGACCTGGGAGACTACAGATGCTATGTCACAGGAGGCC 1322
Db 724 -----ProValGlnAlaP 728
QY 1323 CAGCAATCTGTCCCGCTGGAGGCTGTCAAGAGGATGCTGGGATAAGCCCTTAGGCAC 1382
Db 728 roThrThrSerSerGluThrSerGlnValProThrProSerSerGluSerAsnGlnSerP 748
QY 1383 CAGCTTAGACACTCCCAAGAACCCAGGCCCGCTGTGTAAGATGGCAGATCTGATACCCA 1442
Db 748 roSerGlnAlaProThrProIle-----LeuGluProV 759
QY 1443 TTAGAGCCCGAGAAATCTCTCTTCGTGGATGCCAGTTGCCAGAAACCCACACCCAGCT 1502
Db 759 alHisAlaPro---ThrProAsnSerLysPro-----ValGlnSerProThrProSerS 776
QY 1503 CACACAGCAAAACATGACAGAGGCCAGGCTGACGAAACAGCTGTCCTTCTGGCTG 1562
Db 776 erGluProValSerSerProGluGlnSerGlu-----G 787
QY 1563 TGTGTGAGCTCCAGTAACACCACTATTATTTACCTTTTCCCAAC 1612
Db 787 luValGluAlaProGluProThrProValAsnProSerSerValProSer 803
RESULT 8
QBEB3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BAN>
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-213, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C:Superfamily: human herpesvirus 4 BHLF1 protein
Alignment Scores:
Pred. No.: 1-11e-05 Length: 660
Score: 192.00 Matches: 145
Percent Similarity: 32.28% Conservative: 49

Best Local Similarity: 24.13% Mismatches: 192
Query Match: 6.14% Indels: 215
DB: 1 Gaps: 36
US-09-762-021a-1 (1-1710) x QBEB3 (1-660)
QY 19 CAGCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCA 78
Db 177 ArgProSerGlyProThrGlyGlyArgProAlaAlaProGlyAlaProGlyThrProAla 196
QY 79 GCCAGGCCAGACAGATGGAGGGGCTCTATGCAAGAGCGCTCCCTATGTAGGACGAGC 138
Db 197 Ala-----ProGlyProGlyGlyAla-----AlaValProSerGlyAlaThr 211
QY 139 ACGCTA-----TCTGGAGCGGGGATCTCCCTCAGAACACAGCC-----CACCA 180
Db 212 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProPro 231
QY 181 GAG-----GACCTTAGA----- 192
Db 232 GluArgGlnGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGly 251
QY 193 -----GCACAGCCTCCACCATC 210
Db 252 ProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 271
QY 211 CCCAAGGCCCTGCCACCCACAGTCCCGAGAACCAAGTGCCTTACTCTGCCTCC 270
Db 272 ProArgSerAlaArgAsnPro---GlyCysProArgThrTrpArgArgSerGlyAla 290
QY 271 TCCAAG-----GCGTCTCTTTCCTCCCGAGGACCCAGAGAGGAGCA 312
Db 291 GlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyGlyArg 310
QY 313 GGAAGT-----GCTGAACCATGCTCTAAGGACATTGAGCT 348
Db 311 ProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla 330
QY 349 GTT-----CATGGAAAGCTGGAGAGGCCAGGCAAGACACGAGCAG 390
Db 331 ValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAla 350
QY 391 GAAGAAGAAATTTGGGAAAAAACAAGGACCCAGGAGGTCTCACCCAGGACAGATACAT 450
Db 351 AlaAlaArgLeuProProGluArgGlnGluPro----- 361
QY 451 TCAGTCTTCCAGAGATCAAGTACAGCTTCAA-----CCTCTGGG 492
Db 362 ---ArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThr 380
QY 493 AAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTCGAGCTGTACACAT-----CCT 546
Db 381 ArgSerGlyAlaAlaAlaGln-----ArgThrHisArgArgPro 393
QY 547 CTTCAAGTCCCTGAA-----CTTCATCTCGCCAGGTGCCCTGAGGTGG----- 591
Db 394 ProGlyCysProArgSerAlaArgAsnProGly---CysProArgThrTrpArgArg 412
QY 592 -----CCTAGCAGCCCAAGTGTCTCACCCTCTCCCTCCCTCCCTCAA 630
Db 413 SerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThr 432
QY 631 AGCTATCAACTGCTACAGTCTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGTT 690
Db 433 GlyGlyArgProAla---AlaPro----- 439
QY 691 GGGCCCAAGCTTGA---CCACTAGCC---GGGCCGACTGGACAGCGCATGAGCCCTGCC 744
Db 440 --GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerG 459
QY 745 CTACCAACCCACATTTCTCAGATGACTGGCACTCCAGAGCCTCCAGCAGCACCTT 804
Db 459 lyAlaThrProHisProGlu---ArgGlySerGlyProAlaAspPro-ProAlaAlaAla 477

Qy 805 AGGATACCAGGACCGCTGTTCCCTTCGGCGGGAAGTCATAGTTAGGAGCACCTCACA 864
| | | | |
Db 478 ArgLeuPro-----ProGluArgGlnGluProArgLeuProGlnAspLeuAla 493
| | | | |
Qy 865 CTTTCCTCAGGAGAGACACAAACCATGACCTCAGCCTGGGAGCCCAACTCCAGGCC 924
| | | | |
Db 494 AlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArg 513
| | | | |
Qy 925 CTC-----CAGCCCAACCTGCCAGCCCTGAAATGCAAGTCTTGACGAGTT 978
| | | | |
Db 514 ThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 533
| | | | |
Qy 979 TGAAGTAGGAACCCAGCGGAACCTGACTGTGTT-----CCAGGAGAGAAAGCTGGA 1029
| | | | |
Db 534 TrpArg---ArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArg 552
| | | | |
Qy 1030 GTTCTGGACACACAAAGCGGTGGTGGCTGGTGAAGANTAGGCGGAGCGGCGCTA 1089
| | | | |
Db 553 ProSerGlyProThr-----GlyGlyArgProAlaAla 563
| | | | |
Qy 1090 CATTCCAAGCAACATCCTGGAGCCCTACAGCCGGGACCCTGGGACCCAGGCCAGTCT 1149
| | | | |
Db 564 -----ProGlyAlaPro-----GlyThrProAlaAlaProGlyPro----- 575
| | | | |
Qy 1150 ACCCTCTGGGTTCCCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCACAGACTGGCT 1209
| | | | |
Db 576 -----GlyGlyGlyAlaAla----- 580
| | | | |
Qy 1210 GCAGCAGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGTGCCCTGACGGGAGCCA 1269
| | | | |
Db 581 -----ValProSerGly----- 584
| | | | |
Qy 1270 GCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCACAGAT 1329
| | | | |
Db 595 AlaThrProHisProGluArgGlySer-----GlyProAlaAsp 597
| | | | |
Qy 1330 CCTGTCCCGCTGGAGGCTGTCAGAAGGATGCTGGGATAGCCCTTAGGCACCACTTA 1389
| | | | |
Db 598 Pro-----ProAlaAlaAlaArgLeu 604
| | | | |
Qy 1390 GACACTCCAGAACCCAGGCCCTGATGTCAGATGCGCATGATACCCATTAGAGC 1449
| | | | |
Db 605 ProProGluArgGlnGluProArgLeu----- 613
| | | | |
Qy 1450 CCCGAGATTCCTCTCTGGATCCAGTTTCAGCAAAACCCACACCCACCTCACAG 1509
| | | | |
Db 614 ProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArg----- 631
| | | | |
Qy 1510 CAAAAACAATGGCAGGCCCCAGAGGCTGAA---GCAAAACAGTGTCCCTTCTGGCTGT 1563
| | | | |
Db 632 -----SerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 646
| | | | |
RESULT 9
T03166
probable immediate early protein - alcelaphine herpesvirus 1
C:Species: alcelaphine herpesvirus 1
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Nov-1999
C:Accession: T03166
R:Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. 71, 6517-6525, 1997
A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.
A:Reference number: 214840; MUID: 97404659; PMID: 9261371
A:Accession: T03166
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1300 <ENS>
C:Cross-references: EMBL:AF005370; NID:g2337967; PIDN:AAG58118.1; PID:g2338034
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Alignment Scores:
Pred. No.: 1.39e-05 Length: 1300
Score: 191.00 Matches: 139

Percent Similarity: 29.89% Conservative: 26
Best Local Similarity: 25.18% Mismatches: 244
Query Match: 6.11% Indels: 144
DB: 2 Gaps: 29

us-09-762-021a-1 (1-1710) x T03166 (1-1300)

Qy 13 GAAGACACCGCTGCAGAAAGCTCTGGAGAAAGAGCTGGAGCAAGACACCTCGACTTGGAGG 72
| | | | |
Db 219 GluGlyProLysGlyGluGlyProGluGlyProGluGlyProGluGlyAspSer----- 236
| | | | |
Qy 73 CTTTCAGCAGCCAGGACAGATGAGGGGCTCTATGAAAGAGCCGCTCCCTATGGA 132
| | | | |
Db 237 ProAspGly---ProGlyAlaGlnGluGlyPro---GluGlyLeuGluGlyProGluGly 254
| | | | |
Qy 133 CGAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCA-----CCAGAGAC 186
| | | | |
Db 255 AspGluGlyProGluGlyProGluGlyProGluGlyGluGlyProGluGlyProGluGly 274
| | | | |
Qy 187 CCTAGACACAGCTCCCAACCATCCCAAGGCCCTGCGCAGCCACACAGTGCCTCCGAGA 246
| | | | |
Db 275 ProLysGlyAspSerProAspGlyProGlyAlaGlnGluGlyProGluGly----- 291
| | | | |
Qy 247 ACCAAGTCCCTTACTCTGCTCTCCCAAGCGGTCTCTTCCCTCCGAGGAGCCAGAGAG 306
| | | | |
Db 292 -----ProGlyGlyProAspGlu 297
| | | | |
Qy 307 GGACGAGAAAGTGTGAACCA---TGTCTTAAGGACATTTAGCTGTTTCATGGGAAAGCT 363
| | | | |
Db 298 AspGluGlyProGluGlyProGluGlyProGluGlyGlu---GlyProGluGlyProGlu 316
| | | | |
Qy 364 GGAGAAGGCCCA---GGCAAGACACAGGAGAAAGAAATTTGGGAAAAAACAAGCA 420
| | | | |
Db 317 GlyGluGlyProGluGlyLeuGluGlyProGluGlyGluGlyProGluGlyProGluGly 336
| | | | |
Qy 421 CCA-----GGGAGTCTCTACCCAGGCACA 444
| | | | |
Db 337 ProGluGlyAspSerProAspGlyProAspAlaGlnGluGlyProGluGlyProGlyGly 356
| | | | |
Qy 445 GTACATTGACTGCTCCAGAAAGATCAAGTACAGCTTCAACCTCTGGGAAAGCTGGCCAC 504
| | | | |
Db 357 Pro-----AspGluAspGluGlyProGluGlyProGluGlyProGluGlyGlu 372
| | | | |
Qy 505 CTGGCTGAAGAGACAAAGTGCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAACCT 564
| | | | |
Db 373 GlyProGluGly-----ProGluGly 379
| | | | |
Qy 565 CATCCTGCCAGGTGCTCCTGAGGTGGCTAGCAGCCCAAGTATCTCACCCCTCCTCAC 624
| | | | |
Db 380 ProGluGlyGluGlyProGluGly---ProGluGlyProGluGlyGluGlyProGluGly 398
| | | | |
Qy 625 CCTTAAGCTATCAACCTGCTACAGTCTCTAGCCCTTCTAGCCACCTGAGAGTAACCTTGGAT 684
| | | | |
Db 399 -----LeuGluGlyProGluGlyGluGlyProGluGly 409
| | | | |
Qy 685 GGGGTTGGGCCAG-----CCTGGACCACTAGCCGGGCGGACTGGA 725
| | | | |
Db 410 ProGluGlyProGluGlyAspSerProAspGlyProGlyAlaGlnGluGlyProGluGly 429
| | | | |
Qy 726 CAGCGATGAGCCCTGCTCCCTACCAACCCACATCTCTAGATGACTGGCAACTTCCAGAGC 785
| | | | |
Db 430 -----ProGluGlyProGluGlyGluGlyProGluGlyLeuGluGly 443
| | | | |
Qy 786 CCTCCAGCAAGCACCCTTAGGATACCA-----GGACCTGTTTCCC 827
| | | | |
Db 444 ProGluGlyGluGlyPro-GluGlyProGluGlyProGluGlyGluGlyProGluGlyPr 463
| | | | |
Qy 828 TTCGCG-----GGGAAGTCATAGTTAGGAGGACCTCACACTTCTCAGGAGAGA 881
| | | | |
Db 463 oGluGlyProGluGlyGluGlyProGluArgProGluGlyProGluGlyGluGlyProGlu 483
| | | | |
Qy 882 CACACAACCATGACCTCAGCCTGGGGACCCCACTCCAGGCGCTCCAGCCCAACCTGAT 941
| | | | |

Db 483 uGlyProGluGlyProGluGlyGluGlyProGlu-----G1 495
Qy 942 CCCAGCCAGCCCTGAAATGCAAGTCTGTAGAGTTTGAAGCTAGGAACCCACGGGAAC 1001
Db 495 yProGluGlyProGluArgAspSer-----503
Qy 1002 TGACTGTGTCAGGAGAGAGTGGAGTTCTGGACACACAGACGGTGTGGCTGG 1061
Db 504 -ProAspGlyProGlyAlaGlnGluGlyProGluGlyProGlu-----G1 518
Qy 1062 TGAAGAAATGAGCGGAGCGGCTGACATTCGAAGCAACATCTGGAGCCCTACAGC 1121
Db 518 yProGluGlyGluGlyProGluGlyLeuGluGlyProGluGlyGluGlyProGluGlyPr 538
Qy 1122 CGGGACCCCTGGGACCCAGGCGCCAGTCACCCCTCTCGGTTTCCAATGCTTGCAGCTAGCT 1181
Db 538 oGluGlyProGluGlyGluGlyProGluGlyProGluGlyProGluGly-----554
Qy 1182 CGAGGCCCTGAAGAGGTACAGACTGGCTGCAGGACAGAACTTCTCCACTGCCACGGTGA 1241
Db 555 -GluGlyProGluGlyPro-----GluGlyPr 563
Qy 1242 GGACACTGGGTCCTGACGGGGAGCCAGCTACTTCGGATAGAGACTGGGAGCTACAGA 1301
Db 563 oGluArgAspSerProAspGly---ProGlyAlaGlnGluGlyProGluGlyProGluG1 582
Qy 1302 TGCTATGTCCACAGGAGGCCACCAATCTGTGTC---CCGGCTGGA---GGCTGTACAGAA 1355
Db 582 yProGluGluAspGluGlyProGluGlyProGluGlyProGluGlyGluGlyProGluG1 602
Qy 1356 GGATCTGGGTAAGCCCTTAGGCACACAGCTTAGACACCTCCAAAGAACCCAGGCCCGCT 1415
Db 602 yProGluGlyGluGlyProGluGlyLeu-----GluGlyProG1 615
Qy 1416 GATGCAAGATGCACATCTGATACCATAGAGCCCGCAGAAATTCCTTCTGGATCCCA 1475
Db 615 uGlyAspGluGlyProGluGluProGlu---GlyProGluGlyAspSerProAspGlyPr 634
Qy 1476 GTTTCAGCAAAACCCACACCCAGCCAGCTCACACAGCAAAAACAAATGGACAGCCAGAGGC 1535
Db 634 o---GlyAlaGlnGluValProGluGlyProlys-----GlyProGluG1 648
Qy 1536 TGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGA 1569
Db 648 yGluCysGlnSerGlyProSerSerCysGluGly 659
RESULT 10
TI9361
hypotheical protein cl761.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: TI9361
R:White, S.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19114
A:Accession: TI9361
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-925 <WILL>
A:Cross-references: EMBL:Z78415; PIDN:CAB01670.1; GSPDB:GN000028; CESP:CL761.4
A:Experimental source: clone CL761
C:Genetics:
A:Gene: CESP:CL761.4
A:Map position: X
A:Introns: 40/3; 98/2; 295/1; 443/3; 590/3; 619/3; 691/1; 810/1; 868/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology:
Alignment Scores:
Pred. No.: 1.46e-05 Length: 925
Score: 190.50 Matches: 155
Percent Similarity: 33.09% Conservative: 72
Best Local Similarity: 22.59% Mismatches: 202
Query Match: 6.09% Indels: 257

DB: 2 Gaps: 35
us-09-762-021a-1 (1-1710) x TI9361 (1-925)
Qy 22 CCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGAGACTCGACTTGGAGGCCTTCAGCC 81
Db 46 ProAlaAsnGlyTyrGlyGlyMetProGlyPro-ProAsnIleGlyGlyProGlnPr 65
Qy 82 A-----GGCAGGACAGATGGAGGGCCCTGCTATGAAAG 117
Db 65 oProGlyGlySerGlnGlnLysGlyGlnGlnGlnPheProGlySerGlyAlaAsnMe 85
Qy 118 GCCGCTCCCT-----ATGGAGCAGCAGCCTATCTGAGCCCGGGAT 159
Db 85 LysLeuProGlyTyrAspGlyAsnSerMetGlnAsnAla---TyrMet---ProTyrPr 103
Qy 160 CCTCCAGAAC-----AGCCCCACAGAGGACCTTAGACACAGCCT 201
Db 103 oPro-GlnAsnGlnArgSerGlyGlyGlnAlaProGlnAsnGlyProProAsnTyrA 123
Qy 202 CCCACCATCCCAAGGCCCTGCCACGACACAGCTGCCGAGAACCA-----G 252
Db 123 snSerHisGlnGlnMetPro-----ProAsnAsnGlnTyrGlyGlyV 137
Qy 253 TGCCT-----TTACTCTGCCTCTCCAAAGCGGCTCT---CTTCCCGCAGGA 297
Db 137 alProAspProTyrArgMetTyrProGlyMetGlnGlyProProGlyGlnValProAsnS 157
Qy 298 CCCAGAGGAGGAGGAGGAGTGTGAACCATGTCTTAAGGACATGAGCTGTTCATGGG 357
Db 157 erGlnAsnSerSer-----161
Qy 358 AAAGCTGGAGAGGCC-----AGCAAGACACAGCAGGAAGAAGAA 399
Db 162 --SerGlnGlnArgProProSerGlnAsnAlaAsnGlnGlnArgProAlaSerGln---- 179
Qy 400 ATTTGGGAAAAAACAAGACAGGAGGTCTCCACCAGCAGCAGTACATTTGACTGTCT 459
Db 180 --AlaGlyGlnGlnTyrProThrGlnAlaLeuProAlaHis-----193
Qy 460 CCAGAGATCAAGTACAGCTTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGAGAC 519
Db 194 -----LeuHisGlyThrProThrTyrProGlyMetPro-----GlnA 207
Qy 520 AAGTCCCTCGAGTCGTACATCTCTTCAAGTCCCTGAACCTTCATCTGCCAGGTG 579
Db 207 snAlaProLeuGlnHisTyrGln-----214
Qy 580 CCTGAGGCTGGCTTAGCAGGCCAAGTATCTACCCCTCTCACCCTAAAG-----C 633
Db 215 -----ProGlnTyrAlaAsnGlyThrProProValArgGlyAlaG 229
Qy 634 TATCAACCTGTACAGTCTCTTAAGCCACCTCAGAGTACCTTTGGATGGGTTGGG 693
Db 229 lySerSerAlaPheProLeu-----236
Qy 694 CCCAGCTGGACCACTAGCCGGCGGCGACTGGACAGCGCATGAGCCCTGCCCTACCAAC- 752
Db 237 --GlnProSerLysGlnSerLysProAspGluHisArg-----ProAsnAsnL 252
Qy 753 --CCACATTCATGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCCTTAGGATA 810
Db 252 euSerAsnSerGlnTyrProGlyAsnPheGlyAlaProGlySerSerSerGlyPheAspS 272
Qy 811 CCAGACACCTGTTTCCCTTCGGCGGGAGTATAGTTAGGAGCAGCCTCACACTTCC 870
Db 272 erPheSerAsnGlySerSerGlyTyrGln-----GlyTyrGlyLeuPro----- 286
Qy 871 TCAGGAGAGACACACACCACTGACCCCTCAGCCTGGGACCCCAACTCCAGCCCTCCAG 930
Db 287 -----GlySerGlyThrProGlySerGlnG 295
Qy 931 CCCCAAACCTGCCCGCAGCCAGCCCTGAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAA 990


```

Db 295 lyHisSerThrSerGlyGlnAsnThrSerAlaGlnAsnAlaThrSerGlyLeuThrGly- 314
    :::      |||      :::::      |||||
QY 991 CCACGGGAAGTACTGTGGTCCAGGAGAGAGCTGAGGTTC- 1035
    |||||      |||      |||||:::
Db 315 ProThrGlyAsnProPheGlySerMetAsnGlyGlnGlyGlyAsnHisGlyGlnPheGly 334
    |||||:::
QY 1036 GGACACACGAAGCGGTGGTGGTGAAGAATGAGCGGAGCGGCTACATGCC 1095
    |||      ::|      |||      |||||
Db 335 GlyAsnAspLeuSerGlyValProGlyProSer---GlyGlyTyrGlyGlnMetAsnSer 353
QY 1096 -----AAGCAACATCTCGAGCCCT-ACAGCCGGGAC 1128
    |||      |||||      ::|      |||||      |||
Db 354 SerAsnThrProAsnGlnSerGluArgSerThrProGlyGlnProSerThr-ProGlyTh 373
QY 1129 CCTCGGACCCAGGCGGAGTCCACCTCTCGGTTCATGCTTCGACTAGTCAGGCC 1188
    |||||:::
Db 373 rProGlySerHisGlyThrValGlySerAsnAlaProLeu-----SerHisG1 389
    |||      |||      |||      |||||
QY 1189 TGAAGAGGTCAGACTGGCTGCAGGCAGAGAACTTCCACTGCCAGCGTGAGGACACT 1248
    |||      |||||      ::|      |||||      |||
Db 389 n-LysProSerGln-----GlnGlnGlnSer-MethIAsnHis- 401
QY 1249 TGGGTCCCTGACGGGAGCCAGCTACTTCGCATATAGACCTGGGAGCTACAGATCCTATG 1308
    |||      |||      |||      ::|
Db 402 -----LeuProHisHisGlnTyrAsnGlnAsnLeu 413
QY 1309 TCCACAGGAGGCCCAAGTCTCTCCGGCTGGAGCTGTGCAGAGGATGCTGGGAT 1368
    |||      |||      ::|      |||||      |||
Db 414 Ser-----ProAsnHisGlyAlaSerSerLeuGlySerGlnGlyHisGlySer 430
QY 1369 AAGCCCTTAGGCACAGCTTAGACACCTCCAGAACACGAGCCCGCTGATCAAGATGCC 1428
    |||:::|||||
Db 431 SerProMetGlySerSerLeuMetProLeuAsnGlyGlnTyrProSerMetThrGlnAsn 450
QY 1428 ----- 1428
Db 451 MetGlnSerProAlaSerThrSerMetGluProThrPheLysGluProAlaValProIle 470
QY 1429 -----AGATCTGATACCATTCATTAGAGCCCGCAG 1455
    |||      |||      |||      |||||
Db 471 ArgHisSerProSerGlnMetProThrHisLeuGlnSerProValHisProSerProAsn 490
QY 1456 -----AATTCCCTCTCTGGATCCCGCTGGAGTTGTCAGCAACCCACACCC 1497
    |||::: |||      |||:::
Db 491 GlyAlaProAlaTyrAsnAlaProSerSerLys-----ThrPro 505
QY 1498 CAGCTCACACGCA- 1512
    |||||
Db 506 AspProThrGlnGlnArgProHisSerProThrPheAlaValProThrLeuProAla 525
QY 1513 -----AAACATGGACAGGC 1527
    |||
Db 526 AlaAlaThrLeuAlaGlnAlaPheSerAlaAsnGlnIleSerThrLysProLysThrSer 545
QY 1528 CCAGAGGCTGAA-----GCAACAGTGTCCCTTCTGGCTGTGGAGCTCCCGACGTA 1581
    |||::: |||      |||||
Db 546 ProGlnLysLysHisHisLysAspGlyValProGluProProThrAlaAspThrProPhe 565
QY 1582 ACCACTATTATTATTACTCTTTCCCAACACTGGAGCATTTATG----- 1626
    |||||      |||
Db 566 ThrThrValThrHisTyrGluLeuProAlaAlaMetThrPheLeuArgAspThrLeuHis 585
QY 1627 -----CCTAGGCTGTGCAGAACTCTGTTAGTCCCTCTCTCT 1662
    |||::: |||
Db 586 ValGlyProAsnAspLysValHisProGlnValGluLysHisTyrPheSerArgLysArg 605
QY 1663 TCTCAA 1668
    |||
Db 606 GlnGln 607
RESULT 11
A54849

```

```

collagen alpha 1(VII) chain precursor - human
N:Alternate names: procollagen alpha 1(VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Sep-1999
C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Ultio, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Cross-references: GB:I02870; NID:g987124; PIDN:AAA75438.1; PID:g987125
R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'EPR', 340-475, 'RALSTASHSTLCWRATRHPCNRGSHWTRAAACEPCNRPRASHRAAG', 524-528, '
A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BAA02853.1; PID:94536
A:Experimental source: keratinocyte
A:Note: the authors translated the codon ACC for residues 394 and 397 as Tyr
R:Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.;
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334380; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892, 'E', 894-1439 <PAR>
A:Cross-references: GB:M6158; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915
A:Experimental source: keratinocyte
R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisayanh, P.S.; Cook, M.E.; Wright,
J. Invest. Dermatol. 99, 691-696, 1992
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion p
A:Reference number: I56328; MUID:93107742; PMID:1469284
A:Accession: I56328
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 'EPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>
A:Cross-references: GB:S51236; NID:g262308; PIDN:AAB24637.1; PID:g262309
R:Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, F.
J. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV colla
A:Reference number: A30296; MUID:89139437; PMID:2537292
A:Accession: A30296
A:Molecule type: protein
A:Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-20
A:Note: two reported peptides cannot be reliably located
R:Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collageno
A:Reference number: I48103; MUID:93271985; PMID:8499916
A:Accession: I84686
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 2395-2871, 'S', 2873-2944 <RF2>
A:Cross-references: GB:L06862; NID:g388713; PIDN:AAA89196.1; PID:g388714
R:Christiano, A.M.; Ryyanen, M.; Ultio, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser is
A:Reference number: A55255; MUID:9422477; PMID:8170945
A:Contents: annotation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL7A1; EBR1; EBD1; EB
A:Cross-references: GDB:128750; OMIM:120120
A:Map position: 3p21.3-3p21.3
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermo
A:Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer

```

C:Function:

A:Description: structural component of extracellular polymer associated with anchoring of
C:Superfamily: unassigned collagens; animal kunitz-type proteinase inhibitor homology;
C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
F:36-201/Domain: von Willebrand factor type A repeat homology <VWAL>
F:321-318/Domain: fibronectin type III repeat homology <FN1>
F:327-413/Domain: fibronectin type III repeat homology <FN2>
F:414-502/Domain: fibronectin type III repeat homology <FN3>
F:508-593/Domain: fibronectin type III repeat homology <FN4>
F:598-683/Domain: fibronectin type III repeat homology <FN5>
F:686-771/Domain: fibronectin type III repeat homology <FN6>
F:776-862/Domain: fibronectin type III repeat homology <FN7>
F:864-952/Domain: fibronectin type III repeat homology <FN8>
F:954-1045/Domain: fibronectin type III repeat homology <FN9>
F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWNA2>
F:1170-1172/Region: cell attachment (R-G-D) motif
F:1189-1253/Region: cysteine/proline-rich
F:1254-2783/Region: interrupted helical
F:1334-1336/Region: cell attachment (R-G-D) motif
F:2008-2010/Region: cell attachment (R-G-D) motif
F:2553-2555/Region: cell attachment (R-G-D) motif
F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
F:2876-2929/Domain: animal kunitz-type proteinase inhibitor homology <BPI>
F:337-786,1109/Binding site: carbohydate (Asn) (covalent) #status predicted
F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
F:2625,2631/Binding site: carbohydate (Lys) (covalent) #status experimental
F:2634,2802,2804/Disulfide bonds: interchain #status predicted

Alignment Scores:

Pred. No.:	1-76e-05	Length:	2944
Score:	190.00	Matches:	144
Percent Similarity:	32.3%	Conservative:	37
Best Local Similarity:	25.71%	Mismatches:	222
Query Match:	6.08%	Indels:	158
DB:	2	Gaps:	31

US-09-762-021A-1 (1-1710) x A54849 (1-2944)

QY	1	GGCAGGCGACTGAAGACGAGCTGTGAGAGAGCTGTGAGAGAGCTGGAGCAAGACC	60
Db	1278	GlyArgThrGlyAlaProGlyPro--GlnGlyProProGlySerAlaThrAlaLysGly	1296
QY	61	TCGACTTGGAGCCCTCAGCC-----AGCCAGGACACATGGAGGGGCGCTGCTATGGA	114
Db	1297	GluArgGlyPheProGlyAlaAspGlyArgProGlySerProGlyArgAla-----Gly	1314
QY	115	AAGCGCGCTCCCTATGGAGCGGACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCC	174
Db	1315	AsnProGlyThrProGlyAlaProGlyLeuLysGlySer-----ProGlyLeuProGly	1332
QY	175	GCACAGAGAGCCCTAGACAGACGCTCCACCATCCCAAGGCC---CCTGCCAGGCCA	231
Db	1333	ProArgGlyAspProGlyGluArgGlyProArgGlyProLysGlyGluProGlyAlaPro	1352
QY	232	CACCACTGCCCCAGAACCAAGTGCCTTTACTCTGCCTCTCCCAAGCGGCTCTCTCC	291
Db	1353	GlyGlnValIleGlyGlyGlyProGlyLeuProGlyArgLysGlyAspProGlyPro	1372
QY	292	CGAGGACCCAGAGGAGGACGAGAGT---GCTGAACCATGTCTCAAGGACATTGAGCT	348
Db	1373	SerGlyProProGlyProArgGlyProLeuGlyAspProGlyProArgGlyProProGly	1392
QY	349	GTTTCATGGGAAGCTGGAGAGGCCAGGCAAGACCCAGCAGAGAGAGAAATTTGGGAA	408
Db	1393	LeuProGlyThrAlaMetLysGlyAspLysGlyAspArgGlyGlu-----Arg	1408
QY	409	AAAAACAGGACCGAGGAGGCTCAC-----GluaspGlyArgAsnGly-----Se	1686
Db	1409	GlyProProGlyProGlyGlyGlyGlyIleAlaProGlyGluProGlyLeuProGlyLeu	1428

QY	436	CCAGGCACAGTACATTGACTGCTCCAGAAGATCAAGTACAGCTTCAACCTCTCTGGGAAG	495
Db	1429	ProGlySer-----ProGlyProGlnGlyProValGlyProProGlyLys	1443
QY	496	GCTGCCACCTGGCTGAAGGAGACAAGTGCCTGAGCTGTCGACATCTCTCAAGTC	555
Db	1444	LysGlyGluLysGlyAspSerGluAsp-----GlyAla	1454
QY	556	CCTGAACATTCCTCTGGCCAGGTGCCCTGAGCGTGGCTAGCAGCCCAAGTATCTCACC	615
Db	1455	ProGlyLeu---ProGlyGln-----ProGlySerProGlyGlnGlnGly	1468
QY	616	CCTCTCACCCCTAAAGCTATCAACCTGCTACAGCTCTCTAAGCCCACTCAGAGTAA	675
Db	1469	ProArgGlyPro-----ProGly-----	1474
QY	676	CCTTTGGATGGGTGGGCCAG-----CCTGGACCACTA-----GCC	713
Db	1475	-----AlaIleGlyProLysGlyAspArgGlyPheProGlyProLeuGlyGluAla	1491
QY	714	GGGCGGACTGGACAGCGGATGAGCCCTGCTACCAACCCACATCTCAGATGACTGCG	773
Db	1492	GlyGluLysGlyGluArgGlyProProGlyProAlaGlySerArgGly---LeuProGly	1510
QY	774	AACTTCCAGAGCCCTCCAGCCCAAGCACCCTTAGGATACCAAGCCTCTTCCCTTCCGC	833
Db	1511	ValAlaGlyArgProGlyAlaLysGlyPro-Glu-----GlyProProGlyProThrG	1528
QY	834	GGGGAAGTCATAGTTAGGAGCACCTCACATTTCTCAGGAGAGACACAAACCA---	891
Db	1528	yArgGlnGly-----GluLysGlyGluProGlyArgProG	1540
QY	892	TGACCTCAGCTGGGGACCCCAACTC-----CAGGCCCTCCAGCCCAACACG	941
Db	1540	yaspProAlaValValGlyProAlaValAlaGlyProLysGlyGlyLysGlyAspValG	1560
QY	942	CCCAGCCAGCCCTGAAAATGCAAGTCTTGTACGAGCTTTGAAGCTAGGAACCCACGGA	1001
Db	1560	yProAlaGlyProArgGlyAlaThrGlyValGln-----	1571
QY	1002	TGACTGTGCTCCAGGAGAGAGCTGGAGGTCTGGACCACAGCAAGCGGTGGTCTGCTG	1061
Db	1572	-----GlyGluArgGlyProProGlyLeuValLeuProG	1583
QY	1062	TCAAGAATGAGCGGCGGACGCGCTACATTTCCAAGCAACATCTCTGAGGCCCTACAGC	1121
Db	1583	yasp-----ProGly-----Proly	1588
QY	1122	CGGGACCCCTGGGACCCAGGCCAGTCACTCCTCGGTTCCTCAATGCTTGCATTTAGCT	1181
Db	1588	sGlyAspProGlyAspArgGlyProIleGlyLeuThrGlyArgAlaGlyProProGlyAs	1608
QY	1182	CGAGCCCTGAAGAGGTTCACAGACTGGCTGCAGGCAGAGAA-----CTTCTCCA	1229
Db	1608	pserGlyProProGlyGluLysGlyAspProGlyArgProGlyProProGlyProValG	1628
QY	1230	CTGCACCGGTGAGGACACTTGGTCTC-----CCTGACGGGAGCC	1268
Db	1628	yProArgGlyArgAspGlyGluValGlyLysGlyAspGlyGlyProProGlyAspPr	1648
QY	1269	AGCTACTTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCCAAG	1328
Db	1648	oglyLeuProGlyLysAla---GlyGluArgGlyLeuArgGlyAlaProGlyValThrG	1667
QY	1329	TCTGTCCCGCTGGAGGCTGTCAAGAGATGCTGGGGATAGCCCTTAGGCACCACTT	1388
Db	1667	yProValGlyGluLysGlyAspGlnGlyAspProGly-----	1679
QY	1389	AGACACCTCCAAGAACCGGCCCGCTGTATGCAAGATGCGAGATCTGTATACCATTAGAG	1448
Db	1680	-----GluaspGlyArgAsnGly-----Se	1686

```
QY 1449 CCCCAGAAATTC-----TCTCTGGATCCAGTTTGGCAGCAAAACCCACACC 1496
    ||| :|||
Db 1686 rProGlySerGlyProLysGlyAspArgGlyGluProGlyProGlyProGlyProGly 1706
    :|||
QY 1497 CCAGCTCACACAGCAAAACAATGACAGAGCCAGAGGCTGAAGCAACACAGTGTCCCT 1554
    :|||
Db 1706 yArgLeuValasp-----ThrGlyProGlyAlaArgGluLysGlyGluPro 1721
    :|||
RESULT 12
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118

Alignment Scores:
Pred. NO.: 1.91e-05 Length: 1188
Score: 189.00 Matches: 140
Percent Similarity: 31.56% Conservative: 50
Best Local Similarity: 23.26% Mismatches: 235
Query Match: 6.04% Indels: 177
Db: 2 Gaps: 23

US-09-762-021A-1 (1-1710) x S49915 (1-1188)
QY 79 GCCAGGCCAGGACAG---ATGGAGGGCGCTGCTATGGAAGCGCTCCCTATGGAGCA 135
    ||| |||
Db 396 AlaArgProValaspCysSerLysHisValCysAlaGlyTyrProThrPro---GlyGly 414
    ||| |||
QY 136 GGCACGCTATCTGGA-----GCCGGGGATCCCTCCAGACAGCCCAACAGAGGACCT 189
    ||| |||
Db 415 GlyProProSerSerProValProGlyLysProAlaAlaSerAlaProMetProSerPro 434
    ||| |||
QY 190 AGAGCACAGCTCCACCATCCCAAGGCCCTGCCAGCGCCACACACAGTGCCCGAGAAC 249
    ||| |||
Db 435 HisThrProProAspVal-SerProGluProLeuPro-----GluPr 448
    ||| |||
QY 250 AAGTGCCTTTACTCTGCTCTCCTCAAGCGGCTCTCTCCCGGAGGACCCAGAGAGGA 309
    ||| |||
Db 448 oSerProValProAlaProAlaProMetProMetProThrProHisSerProAlaAs 468
    ||| |||
QY 310 CGAGGAAGTGTGAACCATGTCTCAAGGGACATTGAGCTGTTTCATGGGAAAGCTGGAGA 369
    ||| |||
Db 468 pAspTyrVal--ProProThrPro----- 475
    ||| |||
QY 370 GCCCCAGGCAAGACAGCAGGAAGAAGAAATTTGGGAAAAAACAAGGACGAGGAGG 429
    ||| |||
Db 476 -----ProValProGlyLys 480
    ||| |||
QY 430 TCTCACCCAGGCACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCCT 489
    ||| |||
Db 481 SerProProAlaThr-----SerProSerProGlnValGlnProPro 494
    ||| |||
QY 490 GGGAAAGGTGGCCACCTGGCTGGAAGGAGACAAGTGGCCCTGAGCTCGTACACATCCTCT 549
    ||| |||
Db 495 Ala-----AlaSerThrProProPro 501
    ||| |||
QY 550 CAAGTCCCTGAACCTCATCTCGGCCAGGTGCCCTGAGGCTGGCTAGCAGCCCAAGTGCAT 609
    ||| |||
Db 502 SerLeuValLysLeuSerProGlnAlaProValGlySerProProProProValLys 521
    ||| |||
QY 610 CTCACCCCTCCT-----CACCCCTAAAGCTATCAACCTGCTACAGTCTCTG 654
    ||| |||
Db 522 ThrThrSerProProAlaProIleGlySerProSerProProProProValSerValVal 541
    ||| |||
```

```
QY 655 TCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCG 714
    ||| |||
Db 542 SerProProPro-----ProValLysSerProProPro--ProAlaProValGlyS 558
    ||| |||
QY 715 GCGCGACTGGACAGCGGATGAGCCCTGCGCTTACCAACCCACACTTCTCAGATGACTGCCA 774
    ||| |||
Db 558 erProProProGluLysSerProProProAlaProValAlaSerProProProp 578
    ||| |||
QY 775 ACTTCCAGAGCCCTCCAGCCCAAGCACCTT----- 804
    ||| |||
Db 578 roValLysSerProPro-ProProThrLeuValAlaSerProProProProValLysSer 597
    ||| |||
QY 805 -----AGGATACCAAGCACCTGT 822
    ||| |||
Db 598 ProProProProAlaProValAlaSerProProProValLysSerProProProPro 617
    ||| |||
QY 823 TTCCTTCGCGGGGAAGTCAAGTTAGGGAGCACCTCACACTTCTCTCAGGAGAGAC 882
    ||| |||
Db 618 ThrProValAlaSerProPro-----ProProAlaProValAlaSerSerProPro 634
    ||| |||
QY 883 ACACAACCATGACCTCAGCTGAGCGGACCCCAACTCCAGGCCC-----TCCAG 930
    ||| |||
Db 635 ProMet-LysSerProProProProThrProValSerSerProProProProGluLysSe 654
    ||| |||
QY 931 CCCCAAACTGCCAGCCAGCCCTGAAAAATGCAAGTCTTGTACGAGTTTGAAGTAGGAA 990
    ||| |||
Db 654 rProProProProProProAlaLysSerThrProProProGluLysTyr----- 670
    ||| |||
QY 991 CCCAGGGAGTACTGTGTGTCAGGAGAGAGCTGGAGTTCTGGACCAACAGCAAGCG 1050
    ||| |||
Db 671 -ProThrProThr----- 675
    ||| |||
QY 1051 GTGTGTGCTGGTGAAGATGAGCGGAGCGGCTGACATTCCAAGCAACATCCTG-- 1108
    ||| |||
Db 676 -----SerValLysSerSerProProProGluLysSerLeuProProThrLeu 693
    ||| |||
QY 1109 -----GAGCCCTACAGCGGGGACCCCTGGGACCCAGGGCCA 1146
    ||| |||
Db 693 eProSerProProProGlnGluLysProThrProProSerThrProSerLysProProSe 713
    ||| |||
QY 1147 GTCACCTCTCGG-----GTTCCAATGCTTTCGACTTACGAGGCTGAAGAGGTCAC 1200
    ||| |||
Db 713 rSerProGluLysProSerProProLysGluProValSerSerProProGlnThrProLy 733
    ||| |||
QY 1201 AGACTGGCTGCAGGCAGAGAACTTCTCCACT-----GCCACGGTCAG 1242
    ||| |||
Db 733 sSerSerProProProAlaProValSerSerProProProThrProValSerSerProPr 753
    ||| |||
QY 1243 GACACTTGGTCCCTGACGGGGAGCCAGCTACTTTCGCATAAGACCTGGGAGCTACAGAT 1302
    ||| |||
Db 753 oAlaLeuAlaProValSerSerProProSerValLysSerSerProProProAlaProLe 773
    ||| |||
QY 1303 GCTATGTCCAGAGAGGCCCCCAAGATCCTGTCCTCCGGCTGGAGGCTGTCAAGAGATGCT 1362
    ||| |||
Db 773 userSerProProProAlaProGlnValLysSerSerProProProValGln-----Va 791
    ||| |||
QY 1363 GGGGATAAGCCCTTAGCAGCAGCTTAGACACATCCCAAGAACCCGCGCTCATGCAA 1422
    ||| |||
Db 791 lSerSerProProProAlaProLysSerSerProPro----- 803
    ||| |||
QY 1423 GATGGCAGATCTGTATCCATTTAGAGCC-----CCGAG 1455
    ||| |||
Db 804 -----LeuAlaProValSerSerProProGlnValGluLysThrSerProProPr 820
    ||| |||
QY 1456 AATTCTCTTCTGGATCCAGTTCGAGCAAAACCCACAGCCCTCAGCTACACAGCAAAA 1515
    ||| |||
Db 820 oAlaProLeuSerSerProProLeu-AlaProLysSerSerSerProHisValVal---V 839
    ||| |||
QY 1516 CAATGGACAGGCCAGAGGCTGAAGCAACAGTCTCCTTCTGCTGTGTGGAGCCTCC 1575
    ||| |||
Db 839 alSer-SerProProProValValLysSerSerProProProAlaProValSerSerPro 858
    ||| |||
```



```
QY 1348 TGTCAAGAGGATGCTGGGGATAGCCCTTAGGCACACGCTTAGACACCTCCAAGAACCCAG 1407
Db 505 ---GlnGlnValMetArgProThrProSerProThrGlySerSerGlySerArgSerMet 523
QY 1408 GCCCGCTGATGCAGATGGCAGATCTGATACCCATTAGAGCCCGCAGAAATTCCTCTCT 1467
Db 524 SerProAlaValAlaGlnAsnHisProIleSerArgProAlaSerAsnGlnSerSerSer 543
QY 1468 GGATCCCAGTTT-----GCAGCAAAACCCACACCC----- 1497
Db 544 GlyGlyProMetGlnGlnProProValGlyAlaGlyGlyProProMetProProHis 563
QY 1498 -----CAGCTCACACAGCAAAACCAATGGACAGGC 1527
Db 564 ProGlyMetProGlyGlyProGlnGlnGlnSerGlnGlnGlnAlaSerAsn 583
QY 1528 CCAGAGGCTGAAGCAACAGTCTCCTGCTGCTGTGTGGAGCCTCCCGAGTAACACACC 1587
Db 584 SerAlaSerSerAlaSerAsnSerProGlnGlnThrProProAlaProPro----- 601
QY 1588 TATTTATTACCTTCTCCCAACCTGGAGCATTTATGCTGAGGCTTGTCAGAGATCTG 1647
Db 602 -----ProAsnGlnGly-----MetAsnAsnMet 609
QY 1648 TTCAGTCCCTCTCCT 1662
Db 610 AlaThrProProPro 614
RESULT 14
T29074
Hypothetical protein SC1C2.25c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T29074
Mol. Microbiol. 21, 77-96, 1996
A:Title: A set of ordered cosmid clones and a detailed genetic and physical map for the 8 Mb S
A:Reference number: Z20556; MUID:97000351; PMID:8843436
A:Accession: T29074
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1329 <RED>
A:Cross-references: EMBL:AL031124; NID:e1312893; PID:e1312918; PIDN:CAAL19992.1
C:Genetics:
A>Note: SC1C2.25c
Alignment Scores:
Pred. No.: 3.39e-05 Length: 1329
Score: 185.50 Matches: 139
Percent Similarity: 31.15% Conservative: 51
Best Local Similarity: 22.79% Mismatches: 220
Query Match: 5.93% Indels: 201
Db: 2 Gaps: 27
US-09-762-021A-1 (1-1710) x T29074 (1-1329)
QY 19 CAGCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCA 78
Db 725 GlnProGlyAlaGlyAsnGlyGlyProAlaAlaGlnAlaGlnAlaAlaGlyValAlaAla 744
QY 79 GCCAGCCAGGACAGATGGAGGGGCGCTGCTATGGAAAGCGCGCTCCCTATGGACGAGC 138
Db 745 AlaArgArgGlyGlyGlnGlyGlyGlyAlaLeuAlaGlyGlyGlyAlaPheGlyGlyGly 764
QY 139 AGC-----CTATCTGGAGCCGGGATGCC-----TCCAGAACGCCCCACAGAGAC 186
Db 765 SerGlyGlyGlyAlaLeuGlyAlaGlyAlaProGlyGlyArgLeuGlyAlaGlyGlnGly 784
QY 187 CCTAGAGCACACCTTCCACCATCCCCAAGGCCCTGCCACGCCACACCATGTCGCCGAGA 246
Db 785 ProArgAlaAlaLeuPro----- 790
```

```
QY 247 ACCAAGTGCCTTTACTGCTCTCCCTCAAGGGCGGTCTCTTCCCGGAGGACCCAGAGAG 306
Db 791 -----GlyArgAspAlaGlyGlyArgProGlyAlaProGlyGly 803
QY 307 GGACGAGGAAGTGTGAACCATGTCTTAAGGACACATTAGCTGTTCATGGGAAACCTGGA 366
Db 804 AlaArgGlyProGlnSerProAlaValProGlyGlnGlnAsnArgProAlaProAlaGly 823
QY 367 GAAGCCCGCAGGCAAGACGAGGAGGAAGAAATTTGGGAAAAAACAAGACCCAGG 426
Db 824 AlaGlyAlaGlyAlaGlyGlnGlnAlaProGlyAlaProGlyGlnGlnAlaAlaGly 843
QY 427 AGGTCTCACCCAGCAGCAGTACATTGCTTCCAGAGATCAAGTACAGCTTCAACCT 486
Db 844 MetAsnAlaPro-----GlnGly 849
QY 487 CTTGGGAAGGTGGCCACTGGCTGAAGGAGACAAGTGCCTCGAGCTGTACACATCT 546
Db 850 AsnAlaProAlaAsnAlaSerGlnGlyGlnAsnMetPheGlyAsnArgAlaPro 869
QY 547 CTTCAAGTCCCTCACTCTCGCCAGGTGGCTGAGGCTGGCTAGCAGCCCAAGT 606
Db 870 GlyAlaProProGlnArgLysArgGlyGly-----Asn-GlyAspAlaGlnGln 886
QY 607 GATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTCTACAGTCTCTGTCTAAGCCACC 666
Db 886 yArgArgProGlnLeuProProArgGly----- 895
QY 667 TGAGAGTAACTTTGGATGGGTGGGGCCACCTGGACCATAGCGGGCGGACTGAC 726
Db 896 -----GlyPro-----ArgAlaGlnLeuPr 902
QY 727 AGCGCATGAGCCCTGCCCTACCAACCCACATCTCAGATGAC-----TGCAACT 777
Db 902 oGlyGlyAsnProGlnProArgValProSerTrpSerAspGluAsnAlaGlnProProVa 922
QY 778 TCCAGAGCCCTCCAGCCCAAGCACCTTAGTATACAGGACCCCTGTTTCCCTTCGGCGGG 837
Db 922 lProArgAlaSerLeuAspAlaProArgGlyHisAspGluProAspSerSerArgThrAs 942
QY 838 A---AGTCATAGGTTA-----GGGAGCACCTCACACTTTCCTCAGGA 876
Db 942 pArgThrProArgLeuAspGluArgGlnGlyProGlySerThrAlaGluMetProAlaVa 962
QY 877 GAAGACACACACCATGACCTGACCTGGGACCCCACTCC----- 919
Db 962 l-----ProArgPheGlyGluProGlnSerProAlaAlaThrAlaGln 976
QY 920 -----AGGCCC-----TCCAGCCCAACCT----- 940
Db 976 uPheAlaArgProAspPheAspAlaProAlaProArgArgAspGluSerGlnAspThrGln 996
QY 941 -----GCCAGCCAGCCCTGAAATGCAAGCTTTGACGAGTTTGAAGCTAGGAACCC 993
Db 996 yGlnThrAlaGlnProGlyGlnAsn-----GlnTyArgAlaArgAsnGln 1011
QY 994 ACGGGAACCT----- 1002
Db 1011 uTyrgluAspGlnTyrglyGlnGlnSerGlnTyrglyGlnAspGlnTyrglyGlnAspGln 1031
QY 1003 -----GACTGTGTCTCCAGGAGAGAAGCT----- 1026
Db 1031 nTyrgluProGlyGlnTyrglyGlnAlaGlyProGlyGlnAspGlnTyrglyGlnAspArgProAr 1051
QY 1027 -----GGAGTTCTGGACACACAGCGGTGGTGGTGGTGAAGATAGGCGGACCGA 1082
Db 1051 gPheGlyGlnAspAlaProGlnAsnAspGlnPheAlaArgProGluProAlaAlaProGln 1071
QY 1083 GGG-----CTACATTCCAAGCAACATCTCGA-----GCCCTACACG 1121
Db 1071 nGlnAspGlyGlyAlaPheValArgSerAspValPheGlyGlnProAlaProAsnGln 1091
QY 1122 CGGGGACCCCTGGGACCCAGGCGCCAGTCAACCTCTCGGTTCCTCAATGCTTCGACTAGCT 1181
```

```

Db 1091 yGlnAlaProAlaAspArgPheAlaAlaSerGlnGlyTyrAspAsnGlySerThr 1109
QY 1182 CGAGGCGCTGAAGAGGTGCACAGACTGGCTGCAGCAGCAGAACTTCCTCCTGCTCCACGGTGA 1241
Db 1110 -----GlyGlnHisGlyLeuProGlyArgGln----- 1118
QY 1242 GGACACTGGGTCCCTGACGGGAGCAGCTACTCTCGCATAAG----- 1284
Db 1119 -----AspProAlaValThrGlyGlnPheGluArgProGlnAl 1131
QY 1285 -----ACCTGGGAGCTACAGACTGCTATGTCACAGAGAGGCCCCACCAATCCTGCTCCC 1337
Db 1131 aGlyGlyThrGlnGlyGlyAsnAsp-----PheGlyAlaProArgProAlaPr 1148
QY 1338 GGCTGGAGGCTGTCACAAGAGCTCTGGGATAGCCCTTAGGCACACAGCTTACACACCTC 1397
Db 1148 o-----GlnArgProValProGlnGluProAlaGlyGlnAsnThrAsnGlyTh 1164
QY 1398 CAGAACCAGGCCCG-----GATACC-----CATTAGAGCCCGAGAAATTC 1460
Db 1164 rProAlaGlnArgProGlyAspGlyTrpAlaLeuProProAlaSerGlyProGlyAspG 1184
QY 1428 CAGATCT-----GATACC-----CATTAGAGCCCGAGAAATTC 1460
Db 1184 yArgThrProLeuTyrAspThrLeuGluThrAsnTrpPheHisGlyAspArgGluSerAr 1204
QY 1461 CTCTCTCGATCCACTTTGCAGCAACCCACACCCAGCTCAGCTCAGCAGCAAAACAATG 1520
Db 1204 gAlaProGlnAlaAlaProAlaAlaSerProGluProGlnAlaAlaGlnProGlnThrPr 1224
QY 1521 GACAGGCCAGAGGCTGAAGCAAAACAGT 1548
Db 1224 oAlaAlaProGlnArgProAlaThrSer 1233

RESULT 15
T19140
hypothetical protein C09G5.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19140
R:Palmer, S.
submitted to the EMBL Data Library, November 1994
A:Reference number: Z19080
A:Accession: T19140
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-963 <WIL>
A:Cross-references: EMBL:Z46791; PIDN:CAA86755.1; GSPDB:GN00020; CESP:C09G5.6
A:Experimental source: clone C09G5
C:Genetics:
A:Gene: CESP-C09G5.6
A:Map position: 2
A:Introns: 48/3; 862/3; 898/1

Alignment Scores:
Pred. No.: 3,58e-05 Length: 963
Score: 185.00 Matches: 141
Percent Similarity: 30.94% Conservative: 44
Best Local Similarity: 23.58% Mismatches: 206
Query Match: 5.92% Indels: 208
DB: 2 Gaps: 31

US-09-762-021A-1 (1-1710) x T19140 (1-963)
QY 34 TCTGGAGGAGAGCT-----GGAGCAAGACCTCGACTTGGAGGCT----- 75
Db 154 SerGlyPheArgAlaProProAlaAlaThrSerThrTyrArgProProHisGlySer 173
QY 76 -----TCAGCCAGG-----CCAGACAG 93
Db 174 AsnTyrAspAsnTyrGlyArgGluProAlaSerSerArgArgProTyrProProGlnGln 193
```

```

QY 94 ATGAGGGGGCGCTGCTATGGAAAGCGCGCTCCTATGGAGCAGGACGCTATCTGTGAGCC 153
Db 194 ProProSerThrSerAlaProHisSerSerProAsnAsnArgThrSerLeuTyrAsnPro 213
QY 154 GGGGATCCTCCAGAACACCCACCAGAGAGACCCCTAGAGCA-----CAG 198
Db 214 GlnProProProGlyThrGlyTyrProThrAsnProArgValProTyrAsnProProGln 233
QY 199 CCT-----CCCACTATCC-----AAGCCCTCCACCGCCACAC 234
Db 234 ProAsnTyrThrArgGlnProThrTyrProGluAspAsnArgAlaProTyrLysProThr 253
QY 235 CAGTCCCGAGAACCAAGTCCTTACTCTGCTCCTCCTCAAGCGGCTCCTCTTCCCCCA 294
Db 254 ArgSerProAsnThrProPro-----ProArg 262
QY 295 GCACCCAGAGAGGAGGAGGAAGTCTGAACCATGCTCCTAAGGACATTGAGCTGTTCAT 354
Db 263 GlnProSerGlyTyrAspSerAspGlyGlnThrProProProSerSerProArgIleTyr 282
QY 355 GGGAAAGCTGGAGAGCGCCAGGCAAGAACACGAGGAGAAAGAAATTTGGGAAAAAAA 414
Db 283 -----AsnThr 284
QY 415 CAAGGACACGAGGAGTCTCACCCAGGCACAGTACATTGCTCTCCAGAGATCAAGTA 474
Db 285 ArgArgProAsnAsnHisGlyProGly-----TyrProGluAspGlnVal 299
QY 475 GAGCTTCAACCTCCCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTTGAGCT 534
Db 300 ProThrAlaProPro-----ValProGlyGlnGln----- 309
QY 535 CGTACACATCCTCTCAAGTCCCTGAACCTTCATCTCTGGCCAGGTGCCCTGAGCTGGCT 594
Db 310 ArgValProProThrGlnThrArgAsn---ProProAsnProThrAsnThrArgGlnPro 328
QY 595 AGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGCTACAGCTCTG 654
Db 329 SerArgPro-----ValProProThrSerAspGlyHisIleGluAlaThrThrPro 345
QY 655 TCTAAGCCCACTGAGAGTAACCTTT-----CGATGGGTTGG 692
Db 346 TyrAsnProSerAlaGlnTyrProThrGlyLysArgGlySerHisProGlyPheGlyPro 365
QY 693 GCCCAGCTGGACCACTAGCCGGCGGCGACTGGCAGGCGCATGAGCCCTGCCCTACCAAC 752
Db 366 GlnArgProArgProGlyThrArgProArgGlyAsnProCysAspGlnCysSerAlaGln 385
QY 753 CCACATTTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCCTTAGGATACC 812
Db 386 ProAsnHisCysProSerGly-----ProProGlyProArgGly-ArgProG 401
QY 813 AGGACCTGTTTCCCTTCGGCGGGAAGTCAATAGTTAGGGAGCACCTCACATTCCTTC 872
Db 401 yProProGlyPhePro-----GlyGlnAspGlyProArgGlyLeuArgGlyLeuAsnG 419
QY 873 AGGAGAAGACACACAACTGACCTCAGCTGGGAGCCCACTCAGCCCTCCAGCCCTCCAGCC 932
Db 419 yGlyTyrSerGlyValGln---ProSerSerTyrAspProValIleGlyCysValGln--- 437
QY 933 CCAACCTGCCCGCAGCCCTGAAATATCAAGTCTTGTACGAGTTTGAAGCTAGGAACC 992
Db 438 -----CysProIleGlyProGlyGluArg-----GlyPr 448
QY 993 CAGGGAACTGACTGTGTCAGGAGAGAACTGGAGTTCTGGACCACACAGACCGGT 1052
Db 448 oAspGlyThrProGlyValProGly---GluAspGlyIleAspGlyGluGlnGly----- 465
QY 1053 GTGTGCTGTGAAGATAGGCGGAGCGGCGGTACATTCCAAGCAACATCTCTGGAGC 1112
Db 466 -ValAsnGlyGlnAsp-----GlyGlnProGlyAl 475
QY 1113 CCTACAGCC-----GGGAGCCCTGGGAGCCAGGCGGCGCTCAC 1151
```

```
Db 475 aProGlyAlaProGlyTyrHisGlyMetAsnGlySerProGlyThrProGlyLysProGly 495
QY 1152 CCTCTCGGGTTCACACTTCTGACCTAGCTCGAGGCTGAAGAGGTACACAGACTGGGTGC 1211
Db 495 yLeuProGlyArgAsnGlyGlnSer----- 503
QY 1212 AGCAGAGAACTTCTCCACTGC----- 1233
Db 504 -----CysLysSerIleProGlyProProGlyGlnProGlyValMe 517
QY 1234 -----CACGGTAGGAC----- 1245
Db 517 tGlyValProGlyArgAspGlyAspProGlyThrAspGlyGluHisGlyGlnAspGlyse 537
QY 1246 -----ACTTGGTCCCTGACGGGGAGCCAGC 1271
Db 537 rProGlyIleGlnGlyProProGlyArgAspGlyThrSerGlyProAspGlyGlnProGly 557
QY 1272 TACTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCCAGCAATCC 1331
Db 557 yValSer-----AlaProGlyAlaProGlyThrAspGlyGlyTyrCysPr 572
QY 1332 TGTCCCGCTGGAGCTGTCAAGAGGATGCTGGGGAT----- 1368
Db 572 oCysProLysArgSerSerLysPheAspPheAsnAspAlaTyrAsnAspAspGluLysAr 592
QY 1369 -----AAGCCCTTAGGCACACAGCTTAGACACCTCCAAAGAACCCAGGCCCC 1412
Db 592 gGlyLeuGluGluHisArgProArgGlyTyrAspSerGluArgAlaGluGluProArgPr 612
QY 1413 GCTGATCAAGATGGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTTCTGGATC 1472
Db 612 o---ArgGlnThrValArgThrAsnThrTyr-----AspGluAsnSerGlyAlaGluHi 629
QY 1473 CCAGTTTTCAGCAAC-----CCCACACCCAGCTCACACAGCAAAACAA 1518
Db 629 sGlnArgArgProAsnTyrGluProSerAlaGluValAlaProProArgGln 646
```

Search completed: February 25, 2003, 19:44:33
Job time : 61.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 19:43:04 ; Search time 32 Seconds

(without alignments)
3320.587 Million cell updates/sec

Title: US-09-762-021A-1

Perfect score: 3127

Sequence: 1 ggcagagcactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 313008

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame-n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09762021/runat_24022003_153040_7696/app_query.fasta_1.1863
-DB=PublishedApplications_AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09762021@cgn_1.1.9 @runat_24022003_153040_7696
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2229	71.3	563	10	US-09-998-598-2591
2	912	29.2	179	9	US-09-764-868-757
3	602	19.3	121	9	US-09-764-868-756
4	598	19.1	134	9	US-09-764-868-1135

5	256.5	8.2	280	9	US-09-764-868-1125	Sequence 1125, Ap
6	256.5	8.2	283	9	US-09-764-868-712	Sequence 712, App
7	186.5	6.0	1023	9	US-09-893-519A-14	Sequence 14, Appl
8	171.5	5.5	1274	9	US-10-020-215-2	Sequence 2, Appli
9	164	5.2	2005	10	US-09-735-367B-3	Sequence 3, Appli
10	163	5.2	2063	10	US-09-735-367B-3	Sequence 2, Appli
11	161.5	5.2	846	10	US-09-858-664A-3	Sequence 3, Appli
12	155.5	5.0	1690	10	US-09-788-043C-5	Sequence 5, Appli
13	154.5	4.9	1806	10	US-09-919-497-56	Sequence 56, Appl
14	154	4.9	1070	10	US-09-735-367B-6	Sequence 6, Appli
15	153.5	4.9	941	12	US-10-124-557-14	Sequence 14, Appl
16	153.5	4.9	1022	12	US-10-124-557-84	Sequence 84, Appl
17	153.5	4.9	1038	12	US-10-124-557-74	Sequence 74, Appl
18	153.5	4.9	1149	12	US-10-124-557-58	Sequence 58, Appl
19	153.5	4.9	1140	12	US-10-124-557-104	Sequence 104, App
20	153.5	4.9	1270	12	US-10-124-557-44	Sequence 44, Appl
21	153.5	4.9	1311	12	US-10-124-557-42	Sequence 42, Appl
22	153.5	4.9	1313	12	US-10-124-557-142	Sequence 142, App
23	153.5	4.9	1314	12	US-10-124-557-50	Sequence 50, Appl
24	153.5	4.9	1320	12	US-10-124-557-46	Sequence 46, Appl
25	153.5	4.9	1320	12	US-10-124-557-60	Sequence 60, Appl
26	153.5	4.9	1354	12	US-10-124-557-48	Sequence 48, Appl
27	153.5	4.9	1361	12	US-10-124-557-40	Sequence 40, Appl
28	153.5	4.9	1363	12	US-10-124-557-52	Sequence 52, Appl
29	153.5	4.9	1404	12	US-10-124-557-2	Sequence 2, Appli
30	153.5	4.9	1404	12	US-10-124-557-62	Sequence 62, Appl
31	152.5	4.9	881	10	US-09-816-860A-2	Sequence 2, Appli
32	151	4.8	661	9	US-10-160-865-14	Sequence 14, Appl
33	151	4.8	1907	9	US-09-832-292-39	Sequence 39, Appl
34	150.5	4.8	827	9	US-10-171-384-3	Sequence 3, Appli
35	149	4.8	419	9	US-10-002-344A-237	Sequence 237, App
36	147	4.7	529	10	US-09-861-597-2	Sequence 2, Appli
37	146.5	4.7	775	9	US-09-738-626-3773	Sequence 3773, Ap
38	146	4.7	575	9	US-10-028-072-128	Sequence 128, App
39	146	4.7	575	9	US-10-121-049-128	Sequence 128, App
40	146	4.7	575	9	US-10-123-904-128	Sequence 128, App
41	146	4.7	575	9	US-10-140-470-128	Sequence 128, App
42	146	4.7	575	9	US-10-175-746-128	Sequence 128, App
43	146	4.7	575	9	US-10-176-918-128	Sequence 128, App
44	146	4.7	575	9	US-10-176-921-128	Sequence 128, App
45	146	4.7	575	9	US-10-137-865-128	Sequence 128, App

ALIGNMENTS

RESULT 1
US-09-998-598-2591
; Sequence 2591, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998, 598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2591
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-998-598-2591

Alignment Scores:
Pred. No.: 5,11e-153 Length: 563
Score: 2229.00 Matches: 428
Percent Similarity: 93.45% Conservative: 0
Best Local Similarity: 93.45% Mismatches: 0
Query Match: 71.28% Indels: 30

DB:	10	Gaps:	1
US-09-762-021a-1 (1-1710) x US-09-998-598-2591 (1-563)			
QY	2	CGAGGCGACTGAAGACCGCTCGAAGAGCTCTGAGGAAGAGCTGGACCAAGACCT	61
DB	136	AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluLeuGluGlnArgPro	155
QY	62	CGACTTGAGCGCTTCAGCCAGGCGGACAGATGGAGGGCGCTGCTATGGAAAGCCG	121
DB	156	ArgLeuGlyGlyLeuGlnProGlyClnAspArgTrpArgGlyProAlaMetGluArgPro	175
QY	122	CTCCCTATGGAGCAGGACCGCTATCTGGAGCGGGGATCCCTCCAGAACGCCACCAG	181
DB	176	LeuProMetGluGlnAlaArgTyLeuGluProGlyLeuProGluGlnProHisGln	195
QY	182	AGGACCTTAGAGCAGCGCTCCACCATCCCAAGGCGCCCTGCCACGCCACAGTGCC	241
DB	196	ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla	215
QY	242	CGAGAACAAGTGCCTTACTCTCCCTCAAGGCGGTCTCTTCCCGGAGGACCCA	301
DB	216	ArgGluProSerAlaPheThrLeuProProProArgArgSerSerProGluAspPro	235
QY	302	GAGAGGACGAGGAAGTGTGAACCATGTCTTAAGGACATTTGAGCTGTTTCATGGGAAAG	361
DB	236	GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys	255
QY	362	CTGAGAGGCGCCAGGCAAGACCGAGGAGGAAGAAATTTGGGAAAAAACAAGAC	421
DB	256	LeuGluLysAlaGlnAlaLysThrSerArgLysLysPheGlyLysLysAsnLysAsp	275
QY	422	CAGGAGGTCTCACCCAGGACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTC	481
DB	276	GlnGlyGlyLeuThrGlnAlaGlnTyIleAspCysPheGlnLysIleLysTySerPhe	295
QY	482	AACCTCTCGGAAGCTGGCCACCTGGCTGAAGAGACAAGTGGCCCTGAGCTCGTACAC	541
DB	296	AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis	315
QY	542	ATCCTCTCAAGTCCCTGAAGTTCATCTGCGCAGGTGCGCTGAGGCTGGCTAGCAGCC	601
DB	316	IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAlaAla	335
QY	602	CAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGCTGAAGC	661
DB	336	GlnValIleSerProLeuLeuLeuThrProLysAlaIleAsnLeuLeuInSerCysLeuSer	355
QY	662	CCACCTGAGAGTAACCTTTGGATGGGTGGGCGCCAGCCTGGACACTAGCGGGCGGAC	721
DB	356	ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrThrSerArgAlaAsp	375
QY	722	TGGACAGCGATAGCCCGCTGCGCTACCAACCCACATTCACAGATGACTGCAACTTCCA	781
DB	376	TrpThrGlyAspGluProLeuProTyGlnProThrPheSerAspAspTrpGlnLeuPro	395
QY	782	GAGCCCTCCAGCCCAAGCACCTTAGATACAGGACCCCTGTTTCCCTTCGCGGGGAAGT	841
DB	396	GluProSerSerGlnAlaProLeuGlyTyIleAsnProValSerLeu	411
QY	842	CATAGTTAGGAGCACCTTCACACTTTCCTCAGGAGAAGACACAAACCATGACCCCTCAG	901
DB	411	-----	411
QY	902	CCTGGGACCCCAACTCGAGGCCCTTCAGGCCCAAACTTGGCCAGCCAGCCCTGAAATG	961
DB	412	-----ArgProSerSerProLysProAlaGlnProAlaLeuLysMet	425
QY	962	CAAGTCTTTGACGAGTTGAAGCTAGGAACCCAGCGAACTGACTGTGGTCCAGGGAGAG	1021
DB	426	GlnValLeuTyGluPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGlu	445
QY	1022	AAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGGTGAAGATGAGCGGGGACGG	1081

DB	446	LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg	465
QY	1082	AGGGGCTACATTCACCAAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTGGGACCAG	1141
DB	466	SerGlyTyIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln	485
QY	1142	GGCAGTCCACCTCTCGGGTTCCAATGCTTTCAGCTTCGAGGCTCGAGGCTGAAGAGTCA	1201
DB	486	GlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluValThr	505
QY	1202	GACTGGCTGAGGAGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGTCCCTGACG	1261
DB	506	AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr	525
QY	1262	GGAGCGCAGCTACTTCGCATAAGACCTGGGAGCTGACAGATGCTATGTCACAGGAGGCC	1321
DB	526	GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuLysProGlnGluAla	545
QY	1322	CCAGGAATCCTGCCCGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCCT	1375
DB	546	ProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro	563
RESULT 2			
US-09-764-868-757			
; Sequence 757, Application US/09764868			
; Patent No. US20020168711A1			
; GENERAL INFORMATION:			
; APPLICANT: Rosen et al.			
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies			
; FILE REFERENCE: PT232			
; CURRENT APPLICATION NUMBER: US/09764,868			
; CURRENT FILING DATE: 2001-01-17			
; Prior application data removed - refer to PALM or file wrapper			
; NUMBER OF SEQ ID NOS: 1510			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 757			
; LENGTH: 179			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
US-09-764-868-757			
Alignment Scores:			
Pred. No.:	3,07e-58	Length:	179
Score:	912.00	Matches:	179
Percent Similarity:	92.27%	Conservative:	0
Best Local Similarity:	92.27%	Mismatches:	0
Query Match:	29.17%	Indels:	15
DB:	9	Gaps:	1
US-09-762-021A-1 (1-1710) x US-09-764-868-757 (1-179)			
QY	796	AGCACCTTAGGATACAGGACCTGTTCCCTTCGCGGGGAGTCAATAGTTAGGGAG	855
DB	1	SerThrLeuArgIleProGlyProCysPhePro	11
QY	856	CACCTCACACTTTCCTCAGGAGAGACACACACATGACCTCAGCTGGGACCCCAA	915
DB	12	-----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs	26
QY	916	CTCCAGGCCCTCCAGCCCCAAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTGTACGA	975
DB	26	nSerArgProSerSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyG	46
QY	976	GTTTGAAGCTAGGAACCCAGCGGAAGTGTGGTCCAGGAGAGAGCTGGAGGTTCT	1035
DB	46	uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyLysLeuGluVal	66
QY	1036	GGACCACAGCAAGCGGTGGTGGTGGTGAAGATGAGCGGGGAGCGGCTACATTCC	1095
DB	66	uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyIlePr	86
QY	1096	AAGCAACATCTCGAGCCCTACAGCGGGGACCCCTTGGGACCCAGGCGGACCCCTC	1155

```
|||||
Db 86 oserAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 106
QY 1156 TCGGGTTCCAACTGCTTACAGTTCGAGCGCTGAAGAGGTACACACTGGCTGCAGGC 1215
Db 106 rArgValProMetLeuArgLeuSerArgProGluGluValThrAspTrpLeuGlnAl 126
QY 1216 AGAAACTTCTCCACTGCCACGGTGAGCACACTTGGGTCCCTGACGGGAGCCAGCTACT 1275
Db 126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
QY 1276 TCGCATTAAGACCTGGGAGGTACAGATGCTATGTCCACAGAGGCCCCACGAATCTCTGTC 1335
Db 146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
QY 1336 CCGCTCGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCT 1375
Db 166 rArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 179
RESULT 3
US-09-764-868-756
; Sequence 756, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 756
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-756
Alignment Scores:
Pred. No.: 6,09e-36 Length: 121
Score: 602.00 Matches: 116
Percent Similarity: 98.33% Conservative: 2
Best Local Similarity: 96.67% Mismatches: 2
Query Match: 19.25% Indels: 0
DB: 9 Gaps: 0
US-09-762-021A-1 (1-1710) x US-09-764-868-756 (1-121)
QY 1016 GGAGAGAAGCTGGAGTTCGGACCACAGCAAGCGGTGGTGGTGGTGAAGATGAGCGC 1075
Db 2 GlyAsp***TrpGlnValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAla 21
QY 1076 GGAGGAGCGGTACATTTCCAAGCAACATCTTGAGGCCCTACAGCGGGGCCCTGGG 1135
Db 22 GlyArgSerGlyTyrrilleProSerAsnIleLeuGluProLeuGlnProGlyThrProGly 41
QY 1136 ACCAGGGCCAGTCACCTCTCGGGTTCCAAATGCTTGGACTTACCTGAGCCCTGAAGAG 1195
Db 42 ThrGlnGlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGlu 61
QY 1196 GTCACACACTGGCTGCAGGCGAGAGAAGTCTCCACTGCCAGGTGAGGACACTTGGGTCC 1255
Db 62 ValThrAspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySer 81
QY 1256 CTGACGGGGAGCCAGCTACTTCCATAAGACCTGGGGAGCTACAGATGCTATCTCCACAG 1315
Db 82 LeuThrGlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGln 101
QY 1316 GAGGCCCCACGAATCTGTCTCCCGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCT 1375
```

```
|||||
Db 102 GluAlaProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 121
RESULT 4
US-09-764-868-1135
; Sequence 1135, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1135
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (84)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1135
Alignment Scores:
Pred. No.: 1,21e-35 Length: 134
Score: 598.00 Matches: 111
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 1
Query Match: 19.12% Indels: 0
DB: 9 Gaps: 0
US-09-762-021A-1 (1-1710) x US-09-764-868-1135 (1-134)
QY 827 CTTCCGGCGGAGTCTAGTGTAGGAGGACCTCACACTTCTCCTAGGAGAAGACACAC 886
Db 14 LeuArgArgGlySerHisArgLeuGlySerThrSerHisPheProGlnGluLysThrHis 33
QY 887 AACCATGACCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGGCCCTCCAGGCCCTGCCCAG 946
Db 34 AsnHisAspProGlnProGlyAspProAsnSerArgProSerSerProLysProAlaGln 53
QY 947 CCAGCCCTGAAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAACCCACCGGAACTGACT 1006
Db 54 ProAlaLeuLysMetGlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThr 73
QY 1007 GTGCTCCAGGAGAGAGCTGGAGGTCTTGACCCACAGCAGCGGTGGTGGTGGTGAAG 1066
Db 74 ValValGlnGlyLysLeuGluValLeu***HisSerLysArgTrpTrpLeuValLys 93
QY 1067 AATGAGCGGGAGCGGAGCGGCTACATTCCAAAGCAACATCTCGAGGCCCTACAGCCGGG 1126
Db 94 AsnGluAlaGlyArgSerGlyTyrrilleProSerAsnIleLeuGluProLeuGlnProGly 113
QY 1127 ACCCTCGGGACCCAGGCCAGTCAACCTCTCGGGTT 1162
Db 114 ThrProGlyThrGlnGlyGlnSerProSerArgVal 125
RESULT 5
US-09-764-868-1125
; Sequence 1125, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
```

; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1125
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1125

Alignment Scores:
Pred. No.: 5,81e-11 Length: 280
Score: 256.50 Matches: 76
Percent Similarity: 43.90% Conservative: 32
Best Local Similarity: 30.89% Mismatches: 67
Query Match: 8.20% Indels: 71
DB: 9 Gaps: 10

US-09-762-021A-1 (1-1710) x US-09-764-868-1125 (1-280)

```
Qy 818 CCTGTTTCCCTTCGGCGGGAAGTCATAGTTAGGGAGCACCTCACACACTTTCTCCTCAGGAG 877
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 15 ProValSer-----ArgGlnSerIleArg---AsnSerGlnLysHisSerPro----- 29

Qy 878 AAGACACACACCAATGACCTCAGCCTCGGGGAC-----CCCAACTCCAGGCCCTCC 928
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30 ---ThrSerGluProThrPro---ProGlyAspAlaLeuProProValSerSerProHis 47

Qy 929 AGCCCCAAACCTGCCAGCCCA-----GCCCTGAAATGCAAGTCTTCTACGAG 976
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 48 ThrHisArgGlyTyrglnProThrProAlaMetAlaLysTyrvAllyIleLeutyRasp 67

Qy 977 TTTGAAGCTAGGAACCCAGCGGAAGTCTGCTGTCAGGAGAGAGAGCTGGAGTTCTG 1036
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 68 PheThrAlaArgAsnAlaAsnGluLeuSerValLeuLysaspGluValLeuValLeu 87

Qy 1037 GACCACACAGCGGTGCTGCTGCTGAAGAAATGAGCGGGAGCGGCTACATTCCA 1096
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 88 GluAspGlyArgGlnTrpTrpLysLeuArgSerArgSerGlyGlnAlaGlyTyrvAlPro 107

Qy 1097 AGCAACATCTCTG----- 1108
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 108 CysAsnIleLeuGlyGluAlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGly 127

Qy 1109 -----GAGCCCTACAGCCGGGAGCCCTCGGACC 1138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 128 GlnLysTyrrpGlyProAlaSerProThrHisLysLeuProProSerPheProGlyAsn 147

Qy 1139 CAGGCCAG----- 1147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 148 LysAspGluLeuMetGlnHisMetAspGluValAsnAspGluLeuIleArgLysIleSer 167

Qy 1148 -----TCACCTCTCTCG-----GTTCCAATG----- 1168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 168 AsnIleArgAlaGlnProGlnArgHisPheArgValGluArgSerGlnProValSerGln 187

Qy 1169 ---CTTCGACTTAGCTGAGGCCTGAAGAGTGCACAGACTGGCTGCAGGCAGAGAACTTC 1225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 ProLeuThrTyrgluSerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPhe 207

Qy 1226 TCCACTGCCAGGTGAGACACTGGGTCTGCTGAGGGGAGCCAGCTACTTCGGATAGA 1285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 208 SerProArgIleValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsn 227

Qy 1286 CTGGGGAGCTACAGATCTATGTCCAGGAGGCCCCACCAATCTCTCCGCTGGAG 1345
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 228 LysGluGluLeuLysValCysGlyGluGluGlyValArgValTyrrSerGlnLeuThr 247

Qy 1346 GCTGTGAGAAGGATCTG 1363
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 MetGlnLysAlaPheLeu 253
```

RESULT 6

US-09-764-868-712

; Sequence 712, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 712
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-712

Alignment Scores:
Pred. No.: 5,82e-11 Length: 283
Score: 256.50 Matches: 76
Percent Similarity: 43.90% Conservative: 32
Best Local Similarity: 30.89% Mismatches: 67
Query Match: 8.20% Indels: 71
DB: 9 Gaps: 10

US-09-762-021A-1 (1-1710) x US-09-764-868-712 (1-283)

```
Qy 818 CCTGTTTCCCTTCGGCGGGAAGTCATAGTTAGGGAGCACCTCACACACTTTCTCCTCAGGAG 877
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 18 ProValSer-----ArgGlnSerIleArg---AsnSerGlnLysHisSerPro----- 32

Qy 878 AAGACACACACCAATGACCTCAGCCTCGGGGAC-----CCCAACTCCAGGCCCTCC 928
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 33 ---ThrSerGluProThrPro---ProGlyAspAlaLeuProProValSerSerProHis 50

Qy 929 AGCCCCAAACCTGCCAGCCCA-----GCCCTGAAATGCAAGTCTTGTACGAG 976
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 51 ThrHisArgGlyTyrglnProThrProAlaMetAlaLysTyrvAllyIleLeutyRasp 70

Qy 977 TTTGAAGCTAGGAACCCAGCGGAAGTCTGCTGTCAGGAGAGAGAGCTGGAGTTCTG 1036
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 71 PheThrAlaArgAsnAlaAsnGluLeuSerValLeuLysaspGluValLeuValLeu 90

Qy 1037 GACCACACAGCGGTGCTGCTGCTGAAGAAATGAGCGGGAGCGGCTACATTCCA 1096
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 91 GluAspGlyArgGlnTrpTrpLysLeuArgSerArgSerGlyGlnAlaGlyTyrvAlPro 110

Qy 1097 AGCAACATCTCTG----- 1108
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 111 CysAsnIleLeuGlyGluAlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGly 130

Qy 1109 -----GAGCCCTACAGCCGGGAGCCCTCGGACC 1138
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 131 GlnLysTyrrpGlyProAlaSerProThrHisLysLeuProProSerPheProGlyAsn 150

Qy 1139 CAGGCCAG----- 1147
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 151 LysAspGluLeuMetGlnHisMetAspGluValAsnAspGluLeuIleArgLysIleSer 170

Qy 1148 -----TCACCTCTCTCG-----GTTCCAATG----- 1168
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 171 AsnIleArgAlaGlnProGlnArgHisPheArgValGluArgSerGlnProValSerGln 190

Qy 1169 ---CTTCGACTTAGCTGAGGCCTGAAGAGTGCACAGACTGGCTGCAGGCAGAGAACTTC 1225
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 191 ProLeuThrTyrgluSerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPhe 210

Qy 1226 TCCACTGCCAGGTGAGACACTGGGTCTGCTGAGGGGAGCCAGCTACTTCGGATAGA 1285
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 211 SerProArgIleValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsn 230

Qy 1286 CTGGGGAGCTACAGATCTATGTCCAGGAGGCCCCACCAATCTCTGCTCCGCTGGAG 1345
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Db	113	ProLeuValProAlaGlyProAlaProPro	:	122
Qy	262	TCTGCTCTCCCAAGCGGTCTCTTCCCGCGAGACCAGAGAGGAGGAGGAGTGC		321
Db	123	-----AlaAlaIysLeu-----		136
Qy	322	-----GAACCATGT-----		330
Db	137	CysAlaProValProAlaAlaAlaAlaAlaAlaGlyProGluProAlaProAlaGly		156
Qy	331	-----CCTAAGGACATTGAGCTGTTCTATGGGAAGCTGGAGAGGCCAGGCAAA		381
Db	157	ProAlaLysProAlaGlyProAlaAlaLeuAlaAlaAArgAlaGlyProGlyProGlyPro		176
Qy	382	GACCAGCAGGAAGAAATTTGGGAAAAAACAAAGGACCAAGGAGGTCTCACCCAGGC		441
Db	177	Gly-----ProGlyProGlyProGly-----		185
Qy	442	ACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCTGGGAAGCTGGC		501
Db	186	-----LysProAlaGlyProGlyAla		192
Qy	502	CACCTGGCTGAAGAGACAAAGTGCCTGAGCTGTACATCTCTTCAAGTCCCTGAA		561
Db	193	AlaGlnThrLeuAsnGlySerAlaAlaLeuLeuAsnSerHisAlaAlaAlaProAla		212
Qy	562	CTTCATCTCGCCAGGTGCCCTGAGCTGGCTAGCAGCCCAAGTATCTACCCCT--		618
Db	213	ValSerLeuValAsn-----		226
Qy	619	CCTCACCCCTAAAGCTATCAACCTGTACAGCTCTGTCTAAAGCCCACTGAGAGTAACT		678
Db	227	ProllysPro-----AlaAlaProGlyThrValIleGlnThrPro-----		240
Qy	679	TTGGATGGGTGG-GCCAGCTGTGACCACTAGCC-----GGG		716
Db	241	PheValGlyAlaAlaAlaProAlaProAlaProAlaAlaProSerProAlaAlaProAla		260
Qy	717	CCGACTGGCAGCGATGAGCCCTGCCCTACCACCCACAT-----TCTCAG		764
Db	261	ProAlaAlaProAlaAlaAlaProProProProProAlaProAlaThrLeuAlaArg		280
Qy	765	ATGACTGGCAACTTCCAGAGCCCTCCAGCCAGCACCCCTTAGATACCAAGACCTGTT		824
Db	281	ProGlyHisProAlaGlyProProThrAlaAlaPro-AlaValProProAlaAla		300
Qy	825	CCCTTCGGCGGGAAGTCAATAGTTTAGGGAGCACTCACACTTCTCAGGAGAAGACAC		884
Db	300	aAlaGlnAsnGlyGly-----SerAlaGlyAlaAlaPr		311
Qy	885	ACAACCATGACCTCACCTCGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCTGCC		944
Db	311	oAlaProAlaAlaAlaGlyGlyProAlaGlyValSerGlyGlnProGlyProGlyAl		331
Qy	945	AGCCAGCCCTGAAATGTCAAGTCTTACGAGTTTCAAGCTAGGAACCCACCGGAACCTGA		1004
Db	331	aAlaAlaAlaProAlaProGlyValIysAlaGluSerProIysArgValGlnAl		351
Qy	1005	CTGTGTGCCAGGAGAAAGCT---GGAGGTTCTGGACACAGCAAGCGGTGGTGGCTGG		1061
Db	351	aAlaProProAlaAlaGlnThrLeuAlaAlaSerGlyProAlaSerThrAlaAlaSer-M		371
Qy	1062	TGAAGANTAGCGGGNACGGCGGTACATTCCAAGC---AANCATCTGGAGCCCTTAC		1118
Db	371	etValIleGlyProThrMetGlnGlyAlaLeuProSerProAlaAlaValProProAla		391
Qy	1119	AGCGGGGACCCCT-----GGGACCCAGGCGCACTCACCT		1154
Db	391	laProGlyThrProThrGlyLeuProIysGlyAlaAlaGlyAlaValThrGlnSerLeuS		411
Qy	1155	CTCGGGTTCCAATG-----CTTCGACTTAGCTCGAGCGCTGAAGAGTCA		1195


```
Qy 1070 -----GAGCGGGAGCGGCTACATTCCAAGCAACAT 1104
      :::::
Db 866 oTyAspSerAsnArgValValLeuArgSerGlyLysAspTyrIleAsnAlaSerCy 886
      :::::
Qy 1105 CTTGGAGCCCTACAGCGGGGCCCT-----GGGACCGAGCGGCTACATCCC 1153
      :::::
Db 886 sValgluGlyLeuSerProTyrCysProLeuValAlaThrGlnAlaProLeuPro 905
      :::::
RESULT 9
US-09-735-367B-3
; Sequence 3, Application US/09735367B
; Patent NO. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Calra, Françoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735, 367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2005
; TYPE: PRT
; ORGANISM: mammal
US-09-735-367B-3

Alignment Scores:
Pred. No.: 0.000424 Length: 2005
Score: 164.00 Matches: 131
Percent Similarity: 33.51% Conservativeness: 60
Best Local Similarity: 22.98% Mismatches: 207
Query Match: 5.24% Indels: 173
DB: 10 Gaps: 27

US-09-762-021A-1 (1-1710) x US-09-735-367B-3 (1-2005)
Qy 32 GCTCTGAGGAAGAGCTGGAGCAAGACCTCGAGTTCGGAGCCCTTACGCGCCAGGAC 91
      :::::
Db 897 SerLeuAspAsnSerGlyProLysLeuProGluPheSerAsnArgProGlyTyrPro 916
      :::::
Qy 92 AGATGGAGGGGCTGTATGGGAAGCGCTCCCTATATGGAGCAGCACCTATCTGGAG 151
      :::::
Db 917 -----SerGlnProValGluGlnArgProLeuGlnMetProProGlnLeuMetGln 934
      :::::
Qy 152 CCGGGGATCCCTCCAGAACACGCCACAGAGGACCCCTAGAGCACAGCCCTCCACCATCC 211
      :::::
Db 935 HisValAlaProProGlnProGlnProGlnProGlnProGlnProGlnProGlnGln 954
      :::::
Qy 212 CCAAGGCCCTGCCACGCCAC-----ACCAGTCCCGAGAACCAAGTGCCTTT 232
      :::::
Db 955 GlnGlnProProProProSerGlnProGlnSerGlnGlnGlnGlnGlnGlnGln 974
      :::::
Qy 233 -----ACCAGTCCCGAGAACCAAGTGCCTTT 259
      :::::
Db 975 MetMetMetMetMetMetGlnGlnAspProLysSerValArgLeuProValSerGln 994
      :::::
Qy 260 ACTTGCTCCTCCAGCGGTCTCTTCCCGGAGGACCCAGAGAGGGGACGAGGAAGTG 319
      :::::
Db 995 AsnValHisProProArg-GlyProLeuAsnPro----- 1005
      :::::
Qy 320 CTGAACCATGTCCTAAGGCATTTGACGTTCATGGGAAGCTGGAGAGGCCAGGCA 379
      :::::
Db 1006 -----AspSerGlnArgMetProMetGln 1013
      :::::
Qy 380 AAGACAGCAGGAAGAATTTGGAAAAAACAAGGACGAGGAGGTCTCACCACG 439
      :::::
Db 1013 nGlnSerGly-SerValProValMetValSerLeuGlnGlyProAlaSerValProPro 1033
      :::::
Qy 440 GCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCCT----- 489
      :::::
```

```
Db 1033 er-----ProAspLysGlnArgMetProMetProValAsnThrProL 1047
      :::::
Qy 490 --GGAGAGGCTGGCCACCTGGCTCAAGAGACAGACTGCCCTGAGCTCGTACACATCCTC 547
      :::::
Db 1047 euGlySerAsnSerArgLysMetValTyrGlnGluSerPro-----GlnAsnProS 1064
      :::::
Qy 548 TTCAAGTCCCTGAACTTCATCCTGGCAGGTGCCCTGAGGCT---GGCTAGCAGCCCAA 604
      :::::
Db 1064 erSerSerProLeuAlaGlu-MetAlaSerLeuProGluAlaSerGlySerGluAlaPro 1083
      :::::
Qy 605 GTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGTCTGTCTAAGGCCA 664
      :::::
Db 1084 SerVal-----ProGlyGlyProAsnAsnMetProSerHisValValLeu 1098
      :::::
Qy 665 CTTGAGAGTAACCTTTTGGATGGGGTGGCCACCTGGACCACTACCGCGGCGGCTGG 724
      :::::
Db 1099 ProGlnAsnGlnLeuMetMet----- 1105
      :::::
Qy 725 ACAGCGGATGAGCCCTGCCCTACCAACCCACATCTCAGATGACTGGCAACTTCCAGAG 784
      :::::
Db 1106 ThrGlyProLysProGlyPro-----SerProLeu 1115
      :::::
Qy 785 CCTTCCAGCAAGCACCCCTTAGGATACCAAGACCCCTTTCCCTTCGCGGGGGAAGTCA 844
      :::::
Db 1116 SerAlaThrGlnGlyAlaThrProGlnGlnProProValAsnSerLeuProSerSerHis 1135
      :::::
Qy 845 AGTTGAGGAGCACCTCACACTTTCCT-----CAGGAG 877
      :::::
Db 1136 -----GlyHisHisPheProAsnValAlaAlaProThrGlnThrSerArgPro 1151
      :::::
Qy 878 AAGACACACAACCAT---GACCCCTCAG-----CCTGGGACCCCACTCCAGGCCC 925
      :::::
Db 1152 LysThrProAsnArgAlaSerProArgProTyrTyrProGlnThrProAsnAsnArgPro 1171
      :::::
Qy 926 TCCAGCCCCAAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTGTACAGAGTTTGAAGCT 985
      :::::
Db 1172 ProSerThrGluProSerGluIleSerLeu----- 1181
      :::::
Qy 986 AGNAACCCAGGGAACTGACTGTGTCAGGAGAGAGAGTGGAGGTTCGTGCACACAGC 1045
      :::::
Db 1182 ---SerProGluArgLeu----- 1186
      :::::
Qy 1046 AAGCGGTGGTGGTGAAGAATGAGCGGACGAGCGGTACATTCCAAGCAACATC 1105
      :::::
Db 1187 -----AsnAlaSerIleAlaGlyLeuPheProProGlnIle 1198
      :::::
Qy 1106 CTGAGGCCCTTACAGCGGGACCCCTGGG-----ACCCAGGCGCAG 1147
      :::::
Db 1199 AsnIleProLeuProPro---ArgProAsnLeuAsnArgGlyPheAspGlnGlnGlyLeu 1217
      :::::
Qy 1148 TCACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCTCGAGAGGCTCAGAGAGTGG 1207
      :::::
Db 1218 AsnProThr-----ThrLeuLysAlaIleGlyGlnAlaProSerAsnLeuThr----- 1233
      :::::
Qy 1208 CTGAGGAGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGTCCCTCAG----- 1260
      :::::
Db 1234 MetAsnProSerAsnPheAlaThrProGlnThrHisLysLeuAspSerValValValAsn 1253
      :::::
Qy 1261 ---GGGAGCGCCACTACTTCGCATTAAGACCTTGGGAGCTACAGATGCT----- 1305
      :::::
Db 1254 SerGlyLysGlnSerAsnSerGlyAlaThrLysArgAlaSerProSerAsnSerArgArg 1273
      :::::
Qy 1306 ATGTCCACAGAGGAGGCCCCAGATCTCTGCCGCTGGAGGCTGTCAAGAGGATGCTGGG 1365
      :::::
Db 1274 SerSerProGlySerSerArgLysThrThrProSerProGlyArgGlnAsnSerLysAla 1293
      :::::
Qy 1366 GATAAGCCCTTAGGCACAGCTTAGACACCTCCAAGAACAGGCGCGCTCATCAAGAT 1425
      :::::
Db 1294 ProLys-----LeuThrLeu---AlaSerGlnThrAsnAlaAlaLeuLeuGlnAsn 1309
      :::::
Qy 1426 GCACAGTCTCATATACCCATTAGAGCCCGAGAAATTCCTTCTTGTGATCCCACT---TTGCA 1482
      :::::
```

Db 1310 -ValGluLeu-----ProArgAsnValLeuValSerProThrProLeuAl 1324
Qy 1483 GCAAAACCCACACCCAGCTCACACAGCAAAACAATGGACAGGCCAGAGGCTGAAGCA 1542
Db 1324 aAsnProProValProGlySerPheProAsnAsnSerGlyLeuAsnProGln---AsnSe 1343
Qy 1543 AACAGTGTCCTTGGCTGTG 1564
Db 1343 rThrValSerValAlaLaVal 1350
RESULT 10
US-09-735-367B-2
; Sequence 2, Application US/09735367B
; Patent No. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Calra, Françoise
; APPLICANT: Anonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2063
; TYPE: PRT
; ORGANISM: Human
US-09-735-367B-2
Alignment Scores:
Pred. No.: 0.000504 Length: 2063
Score: 163.00 Matches: 131
Percent Similarity: 33.51% Conservative: 60
Best Local Similarity: 22.98% Mismatches: 207
Query Match: 5.21% Indels: 173
DB: 10 Gaps: 27
US-09-762-021a-1 (1-1710) x US-09-735-367B-2 (1-2063)
Qy 32 GCTCTGGAGGAGAGCTGGAGCAAGACTCTGACTTGGAGCCCTTCAGCCAGGCCAGGAC 91
Db 955 SerLeuAspAsnSerGlyProLysLeuProGluPheSerAsnArgProGlyTyrPro 974
Qy 92 AGATGGAGGGGGCTGCTATGGAAGGCCGCTCCCTATGGAGCAGCCAGCTACTCTGGAG 151
Db 975 -----SerGlnProValGluGluArgProLeuGlnGlnMetProProGlnLeuMetGln 992
Qy 152 CCGGGGATCCCTCCAGAACAGCCACAGAGGACCCCTAGAGCAGCCTCCACCATCC 211
Db 993 HsValAlaProProGlnProGlnProGlnGlnProGlnProGlnLeuProGlnGln 1012
Qy 212 CCAAGGCCCTGCCAGCCAC ----- 232
Db 1013 GlnGlnProProProSerGlnProGlnSerGlnGlnGlnGlnGlnGlnGln 1032
Qy 233 -----ACCACTGCCCGAGAACCAAGTGCCTTT 259
Db 1033 MetMetMetMetLeuMetMetGlnAspProLysSerValArgLeuProValSerGln 1052
Qy 260 ACTGTCCTCTCCAAAGCGGTCTCTTCCCGGAGACCCAGAGGACCCAGAGGACGAGGAGTG 319
Db 1053 AsnValHisProProArg-GlyProLeuAsnPro----- 1063
Qy 320 CTGAACCATTCCTAAGGACATTGAGCTGTTCATGGGAAAGCTGGAGAGGCCAGGCA 379
Db 1064 -----AspSerGlnArgMetProMetG1 1071
Qy 380 AAGACCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 439
Db 1071 nGlnSerGly-SerValProValMetValSerLeuGlnGlyProAlaSerValProPros 1091

Qy 440 GCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCT----- 489
Db 1091 er-----ProAspLysGlnArgMetProMetProValAsnThrProL 1105
Qy 490 --GGGAAGGTGGCCACCTGGCTCAAGGAGACAAGTCCCTCGAGCTCGTACACATCCTC 547
Db 1105 euGlySerAsnSerArgLysMetValTyrGlnGluSerPro-----GlnAsnProS 1122
Qy 548 TTCAAGTCCCTGAACCTTCATCTGGCCAGGTGCCCTGAGGCT---GGCTAGCAGCCCAA 604
Db 1122 erSerProLeuAlaGlu-MetAlaSerLeuProGluAlaSerGlySerGluAlaPro 1141
Qy 605 GTGATCTCACCCCTCCACCCCTAAAGCTATCAACCTCTACAGTCTCTGCTCAAGCCCA 664
Db 1142 SerVal-----ProGlyGlyProAsnAsnMetProSerHisValLeu 1156
Qy 665 CCTGAGAGTAACCTTTGGATGGGTGGGCCCGCCCTGAGGCT---GGCTAGCAGCCCAA 724
Db 1157 ProGlnAsnGlnLeuMetMet----- 1163
Qy 725 ACAGCGATGAGCCCTGCCCTACCAACCCACACATTCTCAGATGACTGGCAACTTCCAGAG 784
Db 1164 ThrGlyProLysProGlyPro-----SerProLeu 1173
Qy 785 CCCTCCAGCCCAAGCACCCTTAGGATACAGGACCCCTGTTCCCTTCGGCGGGGAAGTCAT 844
Db 1174 SerAlaThrGlnGlyAlaThrProGlnGlnProValAsnSerLeuProSerSerHis 1193
Qy 845 AGGTAGGGAGCCTCAGCTTCTCT-----CAGGAG 877
Db 1194 -----GlyHisHisPheProAsnValAlaLaProThrGlnThrSerArgPro 1209
Qy 878 AAGACACACAACCAT---GACCCTCAG-----CCTGGGGACCCCACTCCAGGCC 925
Db 1210 LysThrProAsnArgAlaSerProArgProTyrTyrProGlnThrProAsnAsnArgPro 1229
Qy 926 TCCAGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGCTTGTACGAGTGTGAAGCT 985
Db 1230 ProSerThrGluProSerGluLeuSerLeu----- 1239
Qy 986 AGGAACCCAGCGGAGCTGACTGTGTCCAGGAGAGAGCTGGAGGTCTCGACCACAGC 1045
Db 1240 ---SerProGluArgLeu----- 1244
Qy 1046 AAGCGTGTGGTGGTGAAGAATGAGCGGAGCGGAGCGGTACATTCCCAAGCAACATC 1105
Db 1245 -----AsnAlaSerIleAlaGlyLeuPheProProGlnIle 1256
Qy 1106 CTGAGGCCCTTACAGCCGGGACCCCTGGG-----ACCCAGGCCAG 1147
Db 1257 AsnIleProLeuProPro---ArgProAsnLeuAsnArgGlyPheAspGlnGlnGlyLeu 1275
Qy 1148 TCACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCGCTGAAGAGGTGACAGACTGG 1207
Db 1276 AsnProThr-----ThrLeuLysAlaIleGlyGlnAlaProSerAsnLeuThr 1291
Qy 1208 CTGAGGCAGAGAACTTCTCCACTGCCAGCGTGGAGACACTTGGGTCCCTGAC----- 1260
Db 1292 MetAsnProSerAsnPheAlaThrProGlnThrHisLysLeuAspSerValValAsn 1311
Qy 1261 ---GGGAGCCAGCTACTTTCGATAAGACCTGGGGAGCTACAGATGCT----- 1305
Db 1312 SerGlyLysGlnSerAsnSerGlyAlaThrLysArgAlaSerProSerAsnSerArgArg 1331
Qy 1306 ATGTCCAGAGAGGCCCCACGAATCCTCTCCCGGTGGAGGTGTGACAGAAGATGCTGGG 1365
Db 1332 SerSerProGlySerSerArgLysThrThrProSerProGlyArgGlnAsnSerLysAla 1351
Qy 1366 GATAGCCCTTAGGCCAGCCTTAGACACCTCCAAAGAACCCAGGCCCGCTGATGCAAGAT 1425
Db 1352 ProLys-----LeuThrLeu---AlaSerGlnThrAsnAlaLeuLeuGlnAsn 1367


```
Db 779 ----GInGlyProAlaAlaProSerGluProLysProHisAlaAlaValPheAlaArg 797
Qy 113 -----CCATAGCAGGCCCTCCATC-----TGTCCT 87
Db 798 ValAlaSerProProGlyAlaProGluLysArgValProSerAlaGlyGlyProPro 817
Qy 86 GGCCTGGCTGAAGGCTCCAAAGTCAGAGTCTTTGGTCCAGAGCTCTTCCAGAGGCTTCT 27
Db 818 ValLeuAlaGluLysAlaArgVal-----ProThrValProArgProGly 833
Qy 26 GCAGGCTGGTCTTCA 12
Db 834 SerSerLeuSerSer 838
RESULT 12
US-09-788-043C-5
; Sequence 5, Application US/09788043C
; Patent No. US20020107361A1
; GENERAL INFORMATION:
; APPLICANT: Zuo, Fengrong, Renu
; APPLICANT: Klonowski, Paul
; TITLE OF INVENTION: No. US20020107361A1el Metalloproteases Having
; TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: ROCH-004
; CURRENT APPLICATION NUMBER: US/09/788,043C
; CURRENT FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: human
US-09-788-043C-5
Alignment Scores:
Pred. No.: 0.00167 Length: 1690
Score: 155.50 Matches: 133
Percent Similarity: 30.32% Conservative: 38
Best Local Similarity: 23.58% Mismatches: 176
Query Match: 4.97% Indels: 217
DB: 10 Gaps: 32
US-09-762-021a-1 (1-1710) x US-09-788-043C-5 (1-1690)
Qy 142 CTATCTGGAGCCGGGATCCCTCCAGAACAGCCCAAGAGGAGCCCTAGAGCAGGCT 201
Db 1092 LeuAlaGlyThrGlyAsp-----ArgThrProProHisSerArgProAlaAlaPro 1109
Qy 202 CCCACATCCCAAGGCCCTCCAGGCCACACAGTCCCGAGGACCAAGTTCCTTTAC 261
Db 1110 SerThrGlySerProValProAlaThr-----GluProPro----- 1121
Qy 262 TCTGCTCTCTCAAGCGGTCTCTTCCCGGAGGACCCAGAGGAGGAGGAGTGTCT 321
Db 1122 AlaAlaLysGluGluGlyValLeuGlyProTrpSerPro----- 1134
Qy 322 GAACCATGTCTAAGGACATTGAGCTGTTCATGGGAAGCTGGAGAGGCCAGGCAAA 381
Db 1135 SerProTrpPro-----SerGlnAlaGlyArgSerProProPro 1147
Qy 382 GACCAGCAG-----GAAGACAAATT----- 402
Db 1148 ProSerGluGlnThrProGlyAsnProLeuIleAsnPheLeuProGluGluAspThrPro 1167
Qy 403 -----TGGAAAAAACAAGGACCCAGG 426
Db 1168 IleGlyAlaProAspLeuGlyLeuProSerLeuSerTrpProArgValSerThrAspGly 1187
Qy 427 AGGTCTCACCCAGGCACAGTACATTGACTGTCTCCAGAA----- 465
```

```
Db 1188 LeuGlnThrProAlaAlaThr-----ProGluSerGlnAsnAspPheProVal 1202
Qy 466 -----GATCAAGTACAGCTTCAACCTCCT-----GGGAAGGCTGGCCACCTGGCT 510
Db 1203 GlyLysAspSerGlnSerGlnLeuProProTrpArgAspArgThrAsnGluValPhe 1222
Qy 511 GAAGGAGACAAGTGCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAACCTTCATCCT 570
Db 1223 LysAspAspGluGluProLysGlyArgGlyAlaPro----- 1234
Qy 571 GGCCAGGTGCCTGAGGCTGAGGCTAGCAGCCCAAGTGTATCTACCCCTCTCCACCCCTAA 630
Db 1235 ---HisLeuPro-----ProArgProSerSerThrLeuProPro-LeuSerPro-- 1249
Qy 631 AGCTATCAACCTGCTACAGTCTCTTAAGCCCACTCAGAGTACCTTTGGATGGGCTT 690
Db 1250 ---ValGlySerThrHisSerSerProSerProAspValAlaGluLeuTrpThrGlyGl 1268
Qy 691 GGGCCAGCCTGGACCACTAGCCGGCCGACTGGAGAGCGGATGAGCCCTGCCCTAGCA 750
Db 1268 yThrValAlaTrpGluProAlaLeuLeuGluGlyGlyLeuGlyProValAspSerGluLeuTr 1288
Qy 751 ACCCACATCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATA 810
Db 1288 pProThrValGlyValAlaLaserLeuLeuProProProIle-----AlaProLeu-- 1304
Qy 811 CCAGGACCTGTTCCTTCCTTCGGCGGGGAAGTACATAGTTAGGGAGCACCTCACACTTCC 870
Db 1305 ----- 1305
Qy 871 TCAGGAGAAGACACACAACCACTGACCCCTCAGCTGGGAGCCCAACTCCAGGCCCTCCAG 930
Db 1305 oGluMetLysValArgAspSerSerLeuGluProGly-----ThrProSerPh 1321
Qy 931 CCCAAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAA 990
Db 1321 eProAlaProGlyProGlySerTrpAspLeuGlnThrValAlaValTrpGlyThrPheLe 1341
Qy 991 CCCAGGGAAGTACTGTGTCAGGAGAGAGAGTGGAGGTTCTGGACCACAGCAAGCG 1050
Db 1341 uProThrThrLeuThrGlyLeu----- 1348
Qy 1051 GTGGTGGCTGGTGAAGAAATGAGCGGGGACGAGCGGTACATTTCCAAGCAACATCCTGGA 1110
Db 1349 -----GlyHisMetProGluProAlaLeu-- 1356
Qy 1111 GCCCTACAGCCGGGACCCCTGGGACCCAGGCGCAGTCACTCTCGGTTTCCAATGCT 1170
Db 1357 -----AsnProGlyProLysGlyGlnProGluSerLeuSerProGluVa 1371
Qy 1171 TCGACTTAGCTCGAGGCTGAAGAGGTCACAGACTGG-----CT 1209
Db 1371 iProLeuSerSerArgLeuLeuSerThrProAlaTrpAspSerProAlaAsnSerHisAr 1391
Qy 1210 GCAGGACGACAACTTCTCCACTG---CCACGGTGGAGGACACTTGGGTGCC--- 1256
Db 1391 gValProGlu-ThrGlnProLeuAlaProSerLeuAlaGluAlaGlyProProAlaAspP 1411
Qy 1257 -----TGACGGGGAGCCAGCTACTTCGCATAGAC----- 1286
Db 1411 roLeuValValArgAsnAlaSerTrpGlnAlaGlyAsnTrpSerGluCysSerThrThrc 1431
Qy 1287 -----CTGGGAGCTAC----- 1298
Db 1431 ysGlyLeuGlyAlaValTrpArgProValArgCysSerSerGlyArgAspGluAspCysA 1451
Qy 1299 -----AGATGTATGTCCACAGAGGCCCCACGAAT 1329
Db 1451 laProAlaGlyArgProGlnProAlaArgArgCys-----HisLeuArgPro----- 1466
Qy 1330 CCTGTCCCGGCTGG-----AGCTGTCAAGAGGATGCTGGGGATAG 1371
```



```
Qy 1092 TTCAAGCAACATCCT-----GGAGCCCTTACAGCC-----GGGGA 1127
Db 1410 gGlyleProGlyProValGlyGluGlnGlyLeuProGlyAlaAlaGlyGlnAspGlyPr 1430
Qy 1128 CCCCTGGAGCCAGCCGAGTCACCTCTCGGGTTCCAAATGCTTCGAGTTAGCTCGAGGC 1187
Db 1430 oProGlyProMetGlyProProGlyLeuProGlyLeuLysGlyAspProGlySerLysG1 1450
Qy 1188 CTGAGAGGTACACAGCTGGCTGTCAGGAGAGAACTTCTCCACTCCACGGTGGAGGAC 1247
Db 1450 yGluLysGlyHisProGlyLeuIleGly-----LeuIleGlyProProGlyGluGlnG1 1468
Qy 1248 TTGGGTCTCTGAC---GGGAGCCAGCTACTTCCGATAGAGCTGGGAGCTACAGATGC 1304
Db 1468 yGluLysGlyAspArgGlyLeuProGlyThrGln-----GlySerProGlyAl 1484
Qy 1305 TATGTCCACAGA-----GGCCCCAGCAATCCTGTC-----CGGCTGG 1343
Db 1484 aLysGlyAspGlyIleProGlyProAlaGlyProLeuGlyProProGlyProProG1 1504
Qy 1344 A-----GGCTGTGAGAGGATGCTGGGGATAAG-----CCCTTAGGCACACAG 1385
Db 1504 yLeuProGlyProGlnGlyProLysGlyAsnLysGlySerThrGlyProAlaGlyGlnLy 1524
Qy 1386 CTTAGACACCTCCAAGAACCCAGCCCGCTGATGCAAGATGCGAGATCTGATACCCATTA 1445
Db 1524 sGlyAspSerGlyLeu--ProGlyProProGlyPro-----ProGly 1537
Qy 1446 GAGCCCCGAGAAATCCTCTTCTGGATCCAGTTTCAGCAAAACCCACACCCAGCTCAC 1505
Db 1538 ProProGlyGluValIleGlnProLeuProIleLeuSerSerLysLysThrArgArgHis 1557
Qy 1506 ACA 1508
Db 1558 Thr 1558

RESULT 14
US-09-735-367B-6
; Sequence 6, Application US/09735367B
; Patent No. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Cairra, Francoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: mammal
US-09-735-367B-6

Alignment Scores:
Pred. No.: 0.00193 Length: 1070
Score: 154.00 Matches: 142
Percent Similarity: 31.98% Conservative: 63
Best Local Similarity: 22.15% Mismatches: 239
Query Match: 4.92% Indels: 198
DB: 10 Gaps: 32

US-09-762-021a-1 (1-1710) x US-09-735-367B-6 (1-1070)
Qy 20 ACCCTGCAGAGGCTCTGGAGCAAGAGCTGGGAGCAAGACCTCGACTTGGAGCCCTTCAG 79
Db 134 AlaIleAsnLeuAlaLeuAlaGlnAsnArgSerGlnAspValArgMetAsnGlyProMet 153
```

```
Qy 80 -----CCAGGCCAG 88
Db 154 GlyAlaGlyAsnSerValArgMetGluAlaGlyPheProMetAlaSerGlyProGlyIle 173
Qy 89 GACAGATGGAGGGGGCTGCT-----ATGGAAAGCCGCTCCCTATATGAGCAGCAGCGC 142
Db 174 IleArgMetAsnAsnProAlaThrValMetIleProProGlyGlyAsnValSerSer 193
Qy 143 TATCTGGAGCCGGGATCCCTCCAGAACACGCCCCACAGAGGACCTAGACACAGCCTC 202
Db 194 MetMetAlaProGlyProAsnProGluLeuGln----- 204
Qy 203 CCACCATCCCCAAGGCCCTGCCACGC----- 229
Db 205 ProArgThrProArgProAlaSerGlnSerAspAlaMetAspProLeuLeuSerGlyLeu 224
Qy 230 CACACCTAGTCCCGAGAACCAAGTGCCTTTACTGTGCCTCTCCAAAGCGGTCTCTTCC 289
Db 225 HisIleGlnGlnGlnSerHisProSerGlySerLeuAlaProProHisHisProMetGln 244
Qy 290 CCCGAGGACCCAGAGAGGAGGACGAGGAAGTGTGAACCATGTCTTAAGGACATTCAGCTG 349
Db 245 ProValSerValAsnArgGlnMetAsnProAlaAsn----- 256
Qy 350 TTTATGGAAAGCTGGAGAGGCCACGAC-----AAAGACCAGCAGGAA 393
Db 257 --PheProGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 275
Qy 394 GAAGAAATTTGGGAAAMAAACAGGACCGAGGAGTCTCACCCAGGACGACAGTACATTGA 453
Db 276 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 290
Qy 454 CTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGAAGGCTGGCCACCTGCTGAA 513
Db 291 -----ProGlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 306
Qy 514 GGAGACAAGTCCCTGAGCTCGTACACATCTCTTCAAGTCCCTGAACTTTCATCCTGGC 573
Db 307 -----ThrAlaProThrGlnValPro----- 313
Qy 574 CAGTGCCTCAGCTGCGCTAGCAGCCCAAGTGAAT---CTCACCCCTCTCCACCCCTAA 630
Db 314 --ValProProGlyTrpAsnGlnLeuProSerGlyAlaLeuGlnProPro----- 329
Qy 631 AGCTATCAACCTGCTACAGTCTCTGTAAAGCCACCTCAGAGTACACCTTTGGATGGGTT 690
Db 330 -----ProAla-GlnGlySerLeuGlyThrMetThrAlaAsnGln-GlyTrp----- 344
Qy 691 GGGCCCTGAGCCTAGCCGGGCGGCGGCTGGACAGCGGATGAGCCCTGCCCT----- 746
Db 345 ----LysLysAlaProLeuProGlyProMetGlnGlnGlnGlnGlnGlnGlnGlnGln 363
Qy 746 ----- 746
Db 363 euAlaThrValGlnThrProSerHisProProProTyProPheGlySerGlnGlnA 383
Qy 747 -----ACCAAC---CCACATTTCTCAGATCAGCTGGCAACTTCCAGAGCCCTC 789
Db 383 laSerGlnAlaHisThrAsnPheProGlnMetSerAsnProGlyGlnPheThrAlaPro- 402
Qy 790 CAGCCAAGCACCCCTT-----AGGATACCAGGACCCCTGTTTCCCTTCGGCG 834
Db 403 GlnMetLysSerLeuGlnGlyGlyProSerArgValProThrProLeu----- 418
Qy 835 GGAAGTCATAGTTAGGAGCACCTCACA-----CTTCTCAGGAGAGAC 882
Db 419 -----GlnGlnProHisLeuThrAsnLysSerProAlaSerSerProSerSer 434
Qy 883 ACACAACCATGACCCCTCAGCTGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCTGC 942
Db 435 PheGlnGlnGlnGlySerProAlaSerSerProThrValAsnGlnThrGlnGlnGlnMetGly 454
Qy 943 CCAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGTAGGACCCACCCAGGAACT 1002
```

```

Db 455 ProArgProGlnAsnAsnProLeu- ||| ||| ||| ||| ||| |||
Qy 1003 GAC -----TGTGTCAGGAGAGAA-----GCTGAGGTTCTGGACACACAG 467
Db 468 GlnGlnProValSerSerProGlyArgAsnProMetValGlnGlnGlyAsnValProPro 487
Qy 1045 CAAGCGGTGCTGCTGTGTAAGAA-----TGAGGCGGGAGCGAGCGGCTACAT----- 1092
Db 488 AsnPheMetValMetGlnGlnGlnProProAsnGlnGlyProGlnSerLeuHisProGly 507
Qy 1093 -----TCCAAAGCAATCCTGGAGCCCTACACCGGGGACCCTGGGACCCA 1140
Db 508 LeuGlyMetProGlyArgLeuProProGlyPheSerAlaGly-----GlnAla 524
Qy 1141 GGGCAGTCACCTCGGGTTCATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCAC 1200
Db 525 AsnProAsnPheMetGlnGlnGlnValProSerThr---ThrAlaThrThrProGlyAsn 543
Qy 1201 -----AGACTGGCTGCAGGAGAGAACTTCTCCACTGCCACGTGAGGACACT 1248
Db 544 SerGlyAlaProGlnLeuGlnAlaAsnGlnAsnValGlnHisAlaGlyGlnGlyAla 563
Qy 1249 TGGTCCCTACGGGAGCAGCTACTTCGCATAAGACCTGGGAGCTACAGATGCTATG 1308
Db 564 GlyProProGlnAsnGlnMetGlnValSerHis----- 574
Qy 1309 TCCACAGGAGGCCACAGAACTCTGTCGCCGCTGGA-----GGCTGCAGAGGATGCT 1362
Db 575 -----GlyProProAsnMetGlnProSerLeuMetGlyIleHisGlyAsnMet 591
Qy 1363 GGGGATAAGCCCTTAGCCACAGC-----TTAGACACCTCCAAAGAACCGGCC 1410
Db 592 AsnAsnGlnGlnAlaGlyThrSerGlyValProGlnValAsnLeuSerAsnMetGlnGly 611
Qy 1411 CCGCTGATGCAAGATGCGACATCTGATACCATAGAGCCCGAGAATTCCTCTCTGGA 1470
Db 612 GlnProGlnGln-----GlyProProSerGlnLeuMetGly 623
Qy 1471 TCCAGTTTCAGCAAAACCCACACCC-----CAGCTCACAGCAAAACCAATGGACAGGC 1527
Db 624 MethisGlnGlnIleValProSerGlnGlnMetValGlnGlnGlnGlyThrLeuAsn 643
Qy 1528 CCAGAG-----GCTGAAGCAACACAGTCTGCTCTGCTGCTGTGTGGAGCC 1572
Db 644 ProGlnAsnProMetIleLeuSerArgAlaGlnLeuMetProGlnGlyGlnMetMetVal 663
Qy 1573 TCCCCAGTACACCTATTATTATTACTCTTTCCTCCAAACCTGGAGCATTTATGCCCT 1629
Db 664 AsnProProSerGln-----AsnLeuGlyProSerProGlnArgMetThrPro 679

RESULT 15
US:10-124-557-14
; Sequence 14, Application US/10124557
; Patent No. US20020137894A1
; GENERAL INFORMATION:
; APPLICANT: Turner, Katherine
; Clark, Stephen C.
; Jacobs, Kenneth
; Hewick, Rodney M.
; Gesner, Thomas G.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/124,557
; FILING DATE: 16-Apr-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cserr, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 941 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-10-124-557-14

Alignment Scores:
Pred. No.: 0.00203 Length: 941
Score: 153.50 Matches: 110
Percent Similarity: 34.48% Conservative: 51
Best Local Similarity: 23.55% Mismatches: 231
Query Match: 4.91% Indels: 75
DB: 12 Gaps: 16

US-09-762-021A-1 (1-1710) x US-10-124-557-14 (1-941)
Qy 157 GATCCCTCCAGACAGACCCCA-----CCAGAGGACCTAGAGACACAGCTCC 204
Db 282 GluProAlaProThrAlaProLysLysProAlaProThrThrProLysGluProAlaPro 301
Qy 205 ACCATCCCCAAGCCCTGCCACCCACACAGTCCCGAGAACCAAGTCTTACTCT 264
Db 302 ThrThrProLysGluProAla---Pro-ThrThrThrLysGluProSerProThrThrPr 320
Qy 265 GCCTCCCTCCAAGCGGTCCTTCCCCCGAGGACCCAGAGAGGACGAGGAGTCTGAA 324
Db 320 oLysGluProAlaProThrThrThrLysSerAlaProThrThrThrLysGluProAlaPr 340
Qy 325 CCATGTCTTAAGGACATTGAGCTGTTTCATGCGGAAGCTGGAGAGCCCGGCAAGAC 384
Db 340 oThrThrThrLys-----SerAlaProTh 348
Qy 385 CAGCAGGAAGAAGAAATTTGGGAAAAAACAAGGACGAGGAGGTCTCACCAGGCACA 444
Db 348 rThrProLysGluProSerProThrThrThrLysGluProAlaProThrProLysG 368
Qy 445 GTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTTCAAGTCCCTGAACT 504
Db 368 u-ProAlaProThrThrProLysProAlaProThrThrProLysGluProAlaProT 388
Qy 505 CTGGCTGAAGGAGACAAGTCCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAACT 564
Db 388 hrThrProLysGluProAlaProThrThr-LysLysProAlaProThrAlaProLys 407
Qy 565 CATCTGCGCCAGGTGCCCTGAGGCTGCGCTAGCAGCCCAAGTGTATCTCACCCCTCTCAC 624
Db 408 GluProAla-----ProThrThrProLysGluThrAlaPro-ThrTh 421

```

```
QY 625 CCCTAAAGCTATCAAC----CTGCTACAGTCTCTTAAGCCCA-----CCTGAGAGTAA 675
    ||||| :
Db 421 rProLysLysLeuThrProThrThrProGluLysLeuAlaProThrThrProGluLys-- 440
    ||||| :
QY 676 CCTTTGGATGGGTGGCCGAGCTGGACACCTAGACCGG-----GCCGACTGGACAGG 729
    ||||| :
Db 441 -----ProAlaProThrThrProGluLysLeuAlaProThrThrPr 454
    ||||| :
QY 730 CGATGAGCCCTGCTACCAACCCACACATCTCAGATGACTGGCAACTTCCAGAGCCCTC 789
    :||| ||| ||||| ||| ||| :|||
Db 454 oGluGlu-ProThrThrProThrProGluLysProAlaPro-ThrThrProLysAlaAla 473
    :||| :||| :||| :||| :|||
QY 790 CAGCCCAAGCACCCTTAGGATACCAAGCCCTGTTCCCTCGGGGGGAAGTCATAGGTT 849
    ||||| :||| ||| ||| |||
Db 474 AlaProAsnThrProLysGluProAlaProThrThrProLysGluProAlaProThrThr 493
    ||||| :||| ||| ||| |||
QY 850 AGGGAGCACCTCACATTTCTCAGGAGAAGACACAAACCA---TGACCTCAGCCCTGG 906
    :||| :||| :||| :||| |||
Db 494 ProLysGluProAlaProThrThrProLysGluThrAlaProThrThrProLysGlyThr 513
    :||| :||| :||| :||| |||
QY 907 GGACCCCAACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAATCCAAGT 966
    ||| ||| ||| ||| |||
Db 514 AlaProThrThrLeuLysGluProAlaProThrThrProLysLysProAlaProLysGlu 533
    ||| ||| ||| ||| |||
QY 967 CTTGTACGAGTTTGAAGCTAGGAACCCACGGGAAGT-----GACTGTGGTCCAGGGAGA 1020
    ||| :||| ||||| ||| |||
Db 534 LeuAlaProThrThrThrLysGluProThrSerThrThrSerAspLysProAlaProThr 553
    :||| :||| :||| :||| |||
QY 1021 GAAGCTGAGGTTCTGGACCCAGCAAGCGGTGGTGGTGAAGAATGAGGGGGGACG 1080
    ||||| |||
Db 554 ThrProLysGlyThrAlaProThrThr----- 562
    ||||| |||
QY 1081 GAGCGGTACATTCCAAGCAACATCTCGAGCCCTACAGCCGGGACCCCTGGGACCCA 1140
    ||| ||| ||| ||| :|||
Db 563 -----ProLysGluProAlaProThrThrProLysGluPro 574
    ||||| |||
QY 1141 GGGCCAGTCACCTCTCGGTTTCCCAATGCTTCGACTAGCTCGAG---GCCTGAAGAGGT 1197
    ||| ||| :||| :||| ||| |||
Db 575 AlaProThrThrProLysGlyThrAlaProThrThrLeuLysGluProAlaProThrThr 594
    :||| :||| :||| :||| |||
QY 1198 CACAGCTGGTGCAGGAGCAACTT-----CTCCACTGCCAGGGTGAGGACACTTGG 1251
    :||| :||| :||| :||| |||
Db 595 ProLysLysProAlaProLysGluLeuAlaProThrThrThrLysGlyProThrSerThr 614
    :||| :||| :||| :||| |||
QY 1252 GTCCCTGAGGGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCC 1311
    ||| ||| :||| :||| ||| |||
Db 615 ThrSerAspLysProAlaProThrThrProLysGluThrAlaProThrThr----- 631
    ||| ||| ||| ||| :||| |||
QY 1312 ACAGGAGGCCCCACGAATCTGCTCCGGCTGGAGGCTGTCAAGAGGATGCTGGGATAAG 1371
    ||| ||| ||| ||| :||| |||
Db 632 -----ProLysGluProAlaProThrThrPro---LysLysProAlaProThrThr 647
    ||| ||| ||| ||| :||| |||
QY 1372 CCCTTAGGCACAGCTTACACACTCCAAGAACACAGGCCCGCTGATGCAAGATGGCAGA 1431
    ||| ||||| :|||
Db 648 ProGluThrProProProThrThrSerGluValSerThrProThrThrThrLysGluPro 667
    :||| :||| :||| :||| |||
QY 1432 TCTGATACCCATTAGAGCCCGAGATCTCTTCTGGATCCCGCTTTCAGCAACCCCC 1491
    :||| :||| :||| :||| :|||
Db 668 ThrThrIleHisLysSerProAspGluSerThr-----ProGluLeuSerAlaGluPro 685
    :||| :||| :||| :||| |||
QY 1492 ACACCCCGAG 1500
    ||||| :|||
Db 686 ThrProLys 688
```

Search completed: February 26, 2003, 00:25:13
Job time : 54 secs